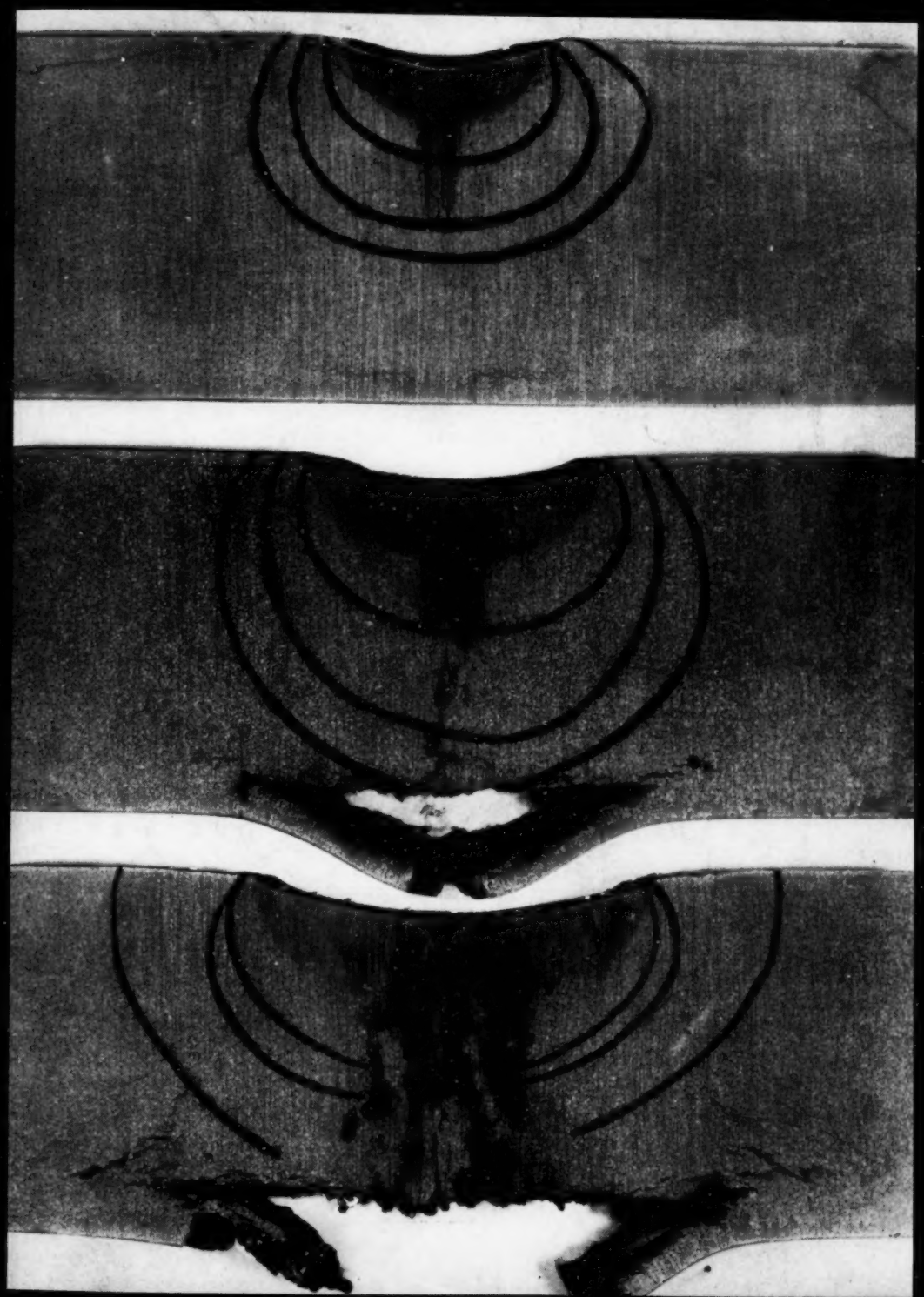
JOURNAL OF APPLIED PHYSICS



JOURNAL OF APPLIED PHYSICS

Vol. 22, No. 9 SEPTEMBER, 1951

In This Issue

	Page
Axially Symmetric Systems for Generating and Measuring Magnetic Fields. Part I Milan Wayne Garrett	1091
Information Theory and Most Efficient Codings for Communication or Memory Devices L. Brillouin	1108
Theory of V-Antennas Ronold King	1111
On the Stress-Function Approaches of Boussinesq and Timpe to the Axisymmetric Problem of Elasticity Theory E. Sternberg, R. A. Eubanks, and M. A. Sadowsky	1121
Offset Wave-Guide Junction as a Reactive Element L. D. Smullin and W. G. Glass	1124
Mathematical Data for Electron Drain on Positive Ion Sources S. E. Rauch	1128
Arcing at Electrical Contacts on Closure. Part II. The Initiation of an Arc L. H. Germer	1133
X-Ray Absorption in a Crystal Set at the Bragg Angle H. N. Campbell	1139
On the Theory of Random Noise. Phenomenological Models. I David Middleton	1143
On the Theory of Random Noise. Phenomenological Models. II David Middleton	1153
Emission of Radiation from Diatomic Gases. II. Experimental Determination of Effective Average Absorption Coefficients of CO S. S. Penner and D. Weber	1164
Evaporation Products of Barium Oxide from Various Base Metals and of Strontium Oxide from Platinum L. T. Aldrich	1168
Allotropic Transformation of Hafnium Pol Duwez	1174
Electrical Charge Storage in Polystyrene Capacitors L. A. Matheson and V. J. Caldecourt	1176
Some Experimental Indications of the Stresses Produced in a Body by an Exploding Charge John S. Rinehart	1178
Reinterpretation of Experiments on Intermetallic Diffusion A. S. Nowick	1182
Generation and Measurement of Low Frequency Random Noise R. R. Bennett and A. S. Fulton	1187
Carbon, Oxygen, and Sulfur Content of Chilean Coppers as Related to Cuprous-Oxide Rectifiers C. C. Hein and W. M. Hickam	1192
Enhanced Emission from Magnetron Cathodes Robert L. Jepsen and Marcel W. Muller	1196
Hollow Cathode for Positive Ion Studies in Cathode-Ray Tubes C. H. Bachman, Harold Eubank, and George Hall	1208
Power Spectrum of a Narrow Band Noise Passed through a Nonlinear Impedance Element John L. McLucas and Richard C. Raymond	1211
Letters to the Editor:	
On Asymmetric Heat Conduction in Copper-Copper Oxide Rectifiers F. Hubbard Horn	1214
Optical Absorption of Barium Oxide with Excess Barium Evan O. Kane	1214
Elastic Vibrations of Almost Cubic Parallelepipeds H. Ekstein and T. Schiffman	1215
Surface Tension of Liquid Metals N. R. Mukherjee	1215
On the Orientations of Zinc Crystals Produced by the Bridgman Method L. M. Slifkin	1216
Note on the Initial Excitation of Linear Systems L. A. Zadeh	1216
Errata: The Electric Tunnel Effect across Thin Insulator Films in Contacts Ragnar Holm	1217
General Formulation for Calculation of Shaped-Beam Antennas A. S. Dunbar	1217
Errata: Application of the Radiation from Fast Electron Beams H. Motz	1217
Announcements:	
Edgar Marburg Lecture Booklet	1218
New Journal	1218
Review of Naturalists' Directory	1218
Conferences	1218
Index to Advertisers	xxxii

BOARD OF EDITORS

Elmer Hutchisson, *Editor*

Earle C. Gregg, *Assistant Editor*

Associate Editors

Howard H. Aiken
Lyle D. Borst
James Hillier
Leonard B. Loeb
W. James Lyons
Eric Reissner
Raymond J. Seeger
Frederick Seitz
Charles S. Smith

Publication of the

American Institute of Physics

OFFICERS

George R. Harrison, *Chairman*
George B. Pegram, *Treasurer*

GOVERNING BOARD

J. W. Beams
C. Paul Boner
Wallace R. Brode
W. F. Fair, Jr.
S. A. Goudsmit
G. P. Harnwell
George R. Harrison
F. V. Hunt
Rudolph Kingslake
Paul Kirkpatrick
F. W. Loomis
J. R. Oppenheimer
George B. Pegram
I. I. Rabi
Duane Roller
Harold K. Schilling
Frederick Seitz
Francis G. Slack
John C. Steinberg

Consultant on Printing

Melvin Loos

ADMINISTRATIVE STAFF

Henry A. Barton, *Director*
Wallace Waterfall, *Executive Secretary*
Theodore Vorburger, *Advertising Manager and Assistant Treasurer*
Ruth F. Bryans, *Publication Manager*
Betty H. Goodfriend, *Assistant Secretary*
Eileen B. Neuberger, *Circulation Manager*
Kathryn Setze *Senior Accountant*

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $10.00. Single copies: $1.00 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 22, Number 9 September, 1951

Axially Symmetric Systems for Generating and Measuring Magnetic Fields. Part I*

MILAN WAYNE GARRETT
Department of Physics, Swarthmore College, Swarthmore, Pennsylvania
(Received August 11, 1950)

This is Part I of a systematic discussion of axially symmetric magnetic fields, both central and remote from the origin, search coils reporting the field or gradient at a single point, and mutual inductances. Here the central uniformity of symmetrical fields and gradients is analyzed by zonal harmonic expansion. Laplace's equation and symmetry restrict these fields to a few types, regardless of the detailed geometry of the generating system. Universal error-contour maps are derived for the central field or gradient in systems having errors of second, fourth, or sixth order, and for hybrid types combining second and fourth. One hybrid has an oblate error field suitable for cloud-chamber and orbital applications.

Source systems include circular filaments, cylindrical or plane circular current sheets, and thick solenoids of rectangular or notched section. Each type of source may be designed to produce any of the field patterns. To this end, source constants derived for the particular source type are combined into a set of over-all coefficients that express the field constants for a complete system. Rapid methods are given for computing the source constants and from them all the field derivatives, using recurrence relations or tables of Legendre functions. In particular, computing time for thick-solenoid fields or gradients is greatly reduced, using a new series with a recurrence formula. The text includes tabular aids and reference formulas, and discusses the rate of convergence of series for central and remote fields, and for mutual inductances.

Special systems briefly described include several infinite series of systems like that which starts with Ampere's loop, the Helmholtz pair, and Maxwell's three-loop system. These have integral numbers of circular filaments from two to infinity. More practical are the thin solenoids. Those that produce a fourth-order (Helmholtz) field by omission of central turns are fully tabulated for all lengths. A short solenoid with double-wound ends and sixth-order error may realize greater uniformity in actual practice than any previously described system. Still greater (theoretical) uniformity is achieved in two eighth-order combinations of a short solenoid with a loop pair; these are described and depicted, with error limits.

INTRODUCTION AND SUMMARY

THIS paper initiates a series that will describe the properties of iron-free wire-wound systems possessing an axis of revolution. These may be primary current-carrying systems or secondary systems for flux measurements (search coils). They consist of coaxial circular turns, isolated or arranged in cylindrical or plane circular sheets, or in multilayer windings of rectangular or notched section. That is, they include filament, disk, and thin or thick solenoid systems. By re-stating the theory in a compact and general mathematical form and by improving methods of calculation, it has been possible to develop a number of new systems. An effort is made to present the essentials so coherently that the reader may dispense with most of the previous literature, which is far from unified. Therefore few historical details are given, and few references cited. Part I includes most of the first three sections of the subjoined outline, with parts of the fourth, but the text does not follow the exact sequence of the summary.

Since this work uses zonal harmonic series throughout, and exploits fully the systematic properties of Legendre's polynomials P_n and P_n', the first section is devoted to these functions. Table I lists their coefficients so that the functions may be recognized wherever found, even though written out in cylindrical coordinates or in terms of the sine as variable. Too often they appear incognito in the literature, expanded into cumbrous polynomials which are evaluated term by term. When properly identified they are readily calculated from tables or often still better from recurrence relations. The latter, though quite familiar to mathematicians, seem to have been little used by physicists in calculating magnetic fields and field coefficients. When the analytical properties of the Legendre functions are fully utilized, mathematical analysis also proceeds more efficiently. Some of these properties are listed, together with tables for finding the zeros of the

* This work was supported in part by a grant from the Research Corporation.

functions, and for computing all-odd or all-even series by a new form of the recurrence relations.

The second section analyzes the magnetic potential and its various derivatives within the central region, whose boundary is a sphere that contains no field source. By a source is meant any discontinuity of the windings, such as an isolated loop or the bounding edge of a solenoid. The theory is clarified and system design made more flexible by immediate separation of the variables, that is by treating independently the coordinates of the field and of its sources. Therefore in this section the sole restrictions are that the potential is finite at the origin and satisfies Laplace's equation, and that there is an axis of revolution. No boundary conditions are imposed by any particular generating system. Such central fields, or field gradients if desired, are limited to a few types. The types are described, with the forms of their equal-error contours, and the shapes of the regions over which they can be made highly uniform. Possible shapes include the highly oblate form which is needed in cloud-chamber and orbital applications. Precisely the same central-field type may arise from a system of circular filaments as from one of thin or thick solenoids or spiral disks, while the error contours of field and of gradient systems are identical in form.

In the third section, the boundary conditions imposed by specific generating systems are derived from a set of numerical constants q_n, calculated from the source coordinates. The calculation requires distinct formulas for each type of source (Eqs. (43)–(54)), but the coefficients when found are used in the same field expansions regardless of the type of source. The complete set q_n defines the central field uniquely, and permits each type of source system to be designed for any required field pattern. The constants of thick solenoids of rectangular section are developed from those of filaments or of current sheets by Taylor expansions carried through the fourth order, and also by a new method of integration that is simpler than any previously published, reducing computing time to a fraction. All expressions for q_n involve the functions P_n', so that tables or recurrence formulas may be employed. To this section is appended a discussion of the rate of convergence of zonal harmonic series, both for fields and for mutual inductances. In view of the new efficiency of calculation, the useful range of such series may be greater than has been supposed.

The fourth section describes new designs for homogeneous fields and gradients. The solenoid forms are simple enough to be wound with high precision; they permit a new order of uniformity, both as single-layer current sheets with overwound ends, and as many-layer solenoids designed for strong fields. For an error set arbitrarily at one part per thousand, the field volume may be expanded fifty-fold over that of an uncorrected solenoid of comparable dimensions. Over a considerable region the uniformity will be limited essentially by the winding errors. Other simple forms include Helmholtz pairs of arbitrary rectangular section, reducing on the one hand to a thin solenoid with equatorial gap, and on the other to a pair of spirally wound current disks. With a simple alteration of axial spacing, the same coils generate the oblate fields already mentioned. The discussion of special fields includes more elaborate systems, some of which may have purely theoretical interest. In this class are the approximations to spherical harmonic current sheets by sets of circular filaments placed on a sphere at positions given by the roots of $P_n=0$ or of $P_n'=0$.

In strict analogy to the analysis of the central field and by a series of nearly identical steps, the remote field will be computed. This comprises the region which lies farther from the origin than any discontinuity of the windings. Here the boundary conditions peculiar to each individual system are described by a set of numerical coefficients p_n analogous to q_n. As before, the finite rectangular section is treated both by series expansion and by integration, and again the resulting expressions involve P_n' and are more easily computed than forms previously given.

The correlation between the remote field of a system and the magnetic flux that it embraces in an external field[1] is further developed and is applied to the design of search-coil systems for use with a fluxmeter to measure the field or its gradient at a single point in an inhomogeneous field. Though the coil has axial symmetry the field need have none. Performance depends upon the number of remote-field error constants p_n that can be suppressed, just as central-field uniformity depends upon suppressing one or more central constants q_n, following q_1 or q_2. The simplest of these coils are evolved from Gray's solenoid; they annul one error term and advance search-coil design as markedly as the Helmholtz pair improved central field uniformity. They have massive homogeneous windings of rectangular section. But if one step in the section profile is allowed, three error terms vanish and the computed residuals are so low that the precision will be limited by winding technique. The discussion is extended to cover so-called "astatic" systems, whose remote field falls off faster than that of a dipole.

The generating or central-field properties are combined with the search or remote-field properties to yield expressions of some generality for the mutual inductance of two axially symmetric systems. These involve the products $q_{n1}p_{n2}$, and they lead to certain new or modified designs of mutual inductance.

It may be noted that in two-dimensional problems[2–4] the cylindrical surface harmonics $\cos n\theta$ or $\sin n\theta$ replace the Legendre functions of the corresponding axially

[1] W. F. Brown and J. H. Sweer, Rev. Sci. Instr. **16**, 276 (1945).
[2] R. H. Lyddane and A. E. Ruark, Rev. Sci. Instr. **10**, 254 (1939).
[3] G. H. Shortley and A. May, J. Appl. Phys. **16**, 841 (1945).
[4] S. K. Haynes and J. W. Wedding, Phys. Rev. **78**, 641 (A) (1950).

symmetric cases. A very complete two-dimensional system can thus be worked out in steps closely following those here outlined, giving fields or gradients of infinitely long straight filaments, plane current strips, or rectangular bundles of finite section. This approach offers the advantage of a systematic guiding analogy, which together with the possible use of recurrence formulas for calculation may facilitate the further development of special two-dimensional fields.

Units, Notation, Sign Conventions

The units are gaussian. The unit generator is a single circular turn carrying current i, whose positive sense bears a right-hand relation to the positive X-axis, as in Fig. 1. When N turns are sufficiently close together, they may be reduced to a single current filament (f) of strength Ni, occupying the mean position. A thin solenoid (s) having N_s' turns per cm of length is treated as a cylindrical current sheet of strength $N_s'i$, while if it constitutes a layer of small radial thickness da in a uniform thick rectangular coil (c) having N'' turns per cm² of section, its strength is $N''ida$. Similarly, the strength of a current disk (d) is $N''idx$ or $N_d'i$, where N_d' is the radial turn density. The current is considered the same in all turns of a system, while the N's are algebraic quantities whose sign must be defined for each type of source. Individual sources are designated by subscripts, j for filaments, k for thin-solenoid end-turns, and l for the bounding edges of disks or thick solenoids.

Referring to Fig. 1, zonal polar coordinates of the jth filament f_j are r_j, α_j, but the cosine u_j appears oftener than α_j. Cylindrical filament coordinates are x_j, a_j. Coordinates of the field point P are r, θ, or x, y.† Note that separate letters are used for source and field cylindrical radii a_j and y, but not for x_j and x. The subscript of x_j will be dropped nevertheless, when in discussing single sources it can cause no confusion. Both $\sin\theta$ and $\sin\alpha_j$ are inherently positive; the latter will usually appear instead of $(1-u_j^2)^{\frac{1}{2}}$. In calculations for solenoids or disks no internal coordinates appear, but the ends of thin solenoids and the edges of disks and thick solenoids are limits of integration which play the role of sources and must be assigned coordinates and algebraic sign. The plus sign belongs to N_j when the jth filament current is positive, to N_k' at the right end of positive and the left end of negative cylindrical sheets, to N_d' at the outer edge of a positive disk, and to N_l' at the outer right-hand vertex of the section of a positive thick rectangular coil. The thick-section perimeter is traversed from this corner with alternation of signs at successive vertices.

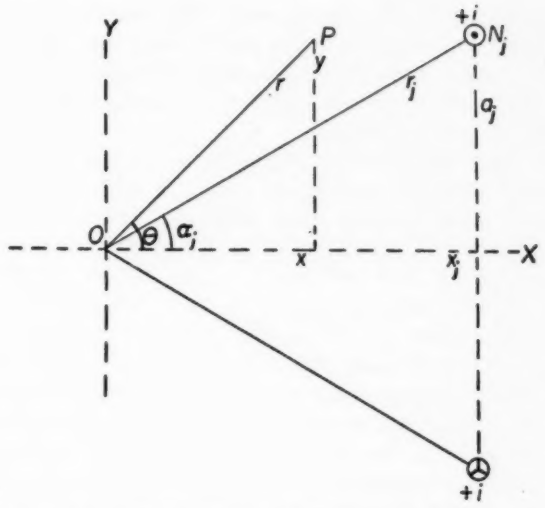


FIG. 1. Meaning of symbols. The diagram shows a positive filament of N_j turns and a central field point P. All j-subscripts become k for thin-solenoid end-turns and l for thick-solenoid edge-turns.

It is convenient to refer the dimensions of all linear coordinates to an arbitrary unit radius r_0, usually the same as some prominent dimension such as the polar or cylindrical radius of some filament or solenoid. This frees the source constants q_n from dimensions; they are pure numbers independent of the absolute scale of any combination.

USE OF THE LEGENDRE FUNCTIONS

All the central field components H_r, H_θ, H_z, H_y, and their derivatives (Eqs. (24)–(30)) are computed as series in which each term is the product of a source constant q_m and a solid zonal harmonic $r^nP_n(\cos\theta)$ or (in the case of y-derivatives) of q_m and $r^nP_n'(\cos\theta)$. The integers m and n may or may not be the same. The factor q_m is a numerical coefficient calculated from source coordinates, while the second factor involves the field coordinates. The remote field series are similar, except that the harmonic factor contains negative powers of r and the source coordinates determine a different set of numerical factors p_m in place of q_m. In every case the source constants involve the Legendre first derivatives. The source constants and the field harmonics should be cleanly parted into separate equations and each complete set evaluated from tables or from recurrence formulas, before multiplying the factors together. Thus the eighth order, for example, is no harder to compute than the third, while the form of calculation preserves the essential similarity between fields of diverse systems having the same symmetry, whose differences must arise within the set of numerical source constants alone. Such similarities have often been obscured by fully expanding the first four or five product polynomials in the series for each specific

† In Fig. 1, P is a central point, since it lies inside a sphere with center at the origin, passing through the nearest source ring. A remote point would lie outside a sphere through the farthest source. To find the field at points between such spheres, add together a central-field and a remote-field component, computed from sources respectively outside and inside the sphere through P. Convergence may be improved in some cases by a shift of origin, while if desired the field contributions from different sources may be superposed even though referred to separate origins.

system, with the source and field constants intermingled, and computing the polynomials term by term. These bulky series come to a dead end unless extended by the user, while the labor of calculation snowballs with successive orders.

Further reduplication of formulas has resulted from writing separate expansions for each shift of origin. This is not necessary when field and source variables are separated and the constants q_n are expressed in the general form with the origin placed arbitrarily as in Fig. 1. The series contain both odd and even terms, but in practice one set is eliminated by choosing the origin in one of three ways. First, if the origin lies in a plane of symmetry of the generating system, the field has corresponding symmetry and the even-numbered terms vanish by Eq. (7), while if the plane of the origin is one of antisymmetry, which mirrors every turn with current reversed, the odd harmonic terms vanish. Second, the component field of each source ring may be referred to an origin at the center of that ring. Then $u_k=0$ and only $P_n'(0)$ appears in the series for q_m and p_m. Third, the field at each point may be referred to an origin on the axis in the plane of the point. Here $\cos\theta=0$ and only $P_n(0)$ or $P_n'(0)$ can occur in the harmonic series. Reference to Eqs. (43) to (54) and (4) or (6) will show in each case whether the odd or the even terms are suppressed.

Since the second and third methods refer the field to multiple origins, only the first gives a coherent description, and it is on that account greatly to be preferred, when the symmetry permits its use.‡ Evidently it is easier to design a homogeneous field when all calculations are based on a single origin at its center, while the pairing of symmetrical elements also reduces the labor. Multiple origins are sometimes desirable or even necessary to improve the rate of convergence of the series. For example, in computing fields close to the axis for a lens-type beta-ray spectrometer, the third method should be convenient. It requires only (25), (28), and the coefficients q_n of Eqs. (24) and (43)–(54). The equations developed in this paper allow complete freedom in the choice of origin, and are so convenient in use that more specialized variations need not be cited.

TABLE I. Coefficients of Legendre functions.

		Polynomials P_n					First derivatives P_n'				
n	M_n	A_n	A_{n-2}	A_{n-4}	A_{n-6}	n	M_{n-1}'	A_{n-1}'	A_{n-3}'	A_{n-5}'	A_{n-7}'
0	1	1				1	1	1			
		1						1			
1	1	1				2	3	1			
	*	1					*	1			
		3	1					5	1		
2	$\frac{1}{2}$	2	1			3	$\frac{3}{2}$	4	1		
	(−)	3	2				(−)	5	4		
		5	3					7	3		
3	$\frac{1}{2}$	2	3			4	$\frac{5}{2}$	4	3		
	*(−)	5	2				*(−)	7	4		
		35	30	3				21	14	1	
4	$\frac{1}{8}$	8	24	3		5	$\frac{15}{8}$	8	12	1	
		35	40	8				21	28	8	
		63	70	15				33	30	5	
5	$\frac{1}{8}$	8	40	15		6	$\frac{21}{8}$	8	20	5	
	*	63	56	8			*	33	36	8	
		231	315	105	5			429	495	135	5
6	$\frac{1}{16}$	16	120	90	5	7	$\frac{7}{16}$	64	240	120	5
	(−)	231	378	168	16		(−)	429	792	432	64
		429	693	315	35			715	1001	385	35
7	$\frac{1}{16}$	16	168	210	35	8	$\frac{1}{16}$	64	336	280	35
	(−)*	429	594	216	16		(−)*	715	1144	528	64

Power series in $u=\cos\theta$ (top rows of each set), in x/r, y/r (middle rows), and in $\sin\theta$ (bottom rows).

* In the case of the sine series, multiply M_n or M_{n-1}' by $u/\sin\theta$ when * occurs, and by -1 when (−) occurs.

Coefficients of Legendre Functions

Usually, since a recurrence relation will be used for magnetic field series beyond the second or at most the third term, there is no need for further expansion. But to correlate with earlier work, Table I lists the first seven or eight internal coefficients of the polynomials $P_n(u)$ and of $P_n'(u)$, their first derivatives with respect to u. Those in the top line of each set of three are coefficients of the familiar cosine series:

$$P_n(u)=M_n(A_nu^n-A_{n-2}u^{n-2}+A_{n-4}u^{n-4}-\cdots)$$

$$P_n'(u)=M_{n-1}'(A_{n-1}'u^{n-1}-A_{n-3}'u^{n-3}+A_{n-5}'u^{n-5}-\cdots)$$

while the middle lines correspond to power series in x/r, y/r, and the bottom lines to series in $\sin\theta$, thus:

$$P_0=1 \quad P_1=u\equiv x/r$$

$$P_2=\tfrac{1}{2}(3u^2-1)=\tfrac{1}{2}r^{-2}(2x^2-y^2)=-\tfrac{1}{2}(3\sin^2\theta-2)$$

$$P_3=\tfrac{1}{2}u(5u^2-3)=\tfrac{1}{2}r^{-3}x(2x^2-3y^2)=-\tfrac{1}{2}u(5\sin^2\theta-2)$$

$$P_4=\tfrac{1}{8}(35u^4-30u^2+3)=\tfrac{1}{8}r^{-4}(8x^4-24x^2y^2+3y^4)$$

$$=\tfrac{1}{8}(35\sin^4\theta-40\sin^2\theta+8). \quad (1)$$

Though Curtis,[5] Dwight,[6] and Koch[7] among others have employed series of sines, the x, y series (cylindrical coordinates) occur most commonly. There is good evidence that most users have recognized Legendre's functions only when written as power series in u. Also a number of published coefficients are in error, some by factors of ten or more, and in one case -105!

‡ Andrew Gray made some use of this method, but later workers, except McKeehan, have seldom employed it.

[5] H. L. Curtis, *Electrical Measurements* (McGraw-Hill Book Company, Inc., New York, 1937), p. 144.

[6] H. B. Dwight, *Electrical Coils and Conductors* (McGraw-Hill Book Company, Inc., New York, 1945), p. 234.

[7] H. W. Koch, J. Appl. Phys. **21**, 387 (1950). Here, on substituting $1/\sin\theta$ for ξ in Eq. (27) of Koch's paper, the long expression may be reduced to a harmonic series in $r^{-(n+1)}P_n(\cos\theta)$ with coefficients containing $1/(n+2)P_n'(0)$. The zero arises from the restricted choice of origin. A later paper of this series will give the unrestricted expression, but in Koch's series the last three harmonics are in error by factors, respectively, 4/3, 1/4, and 1/25.

Many standard texts develop the properties of Legendre's functions,[8] but as the writer has found it difficult to locate the relations most often needed, a few are listed below. Various additional relations among products of these functions by powers of r and of $\sin\theta$ permit whole series to be integrated or differentiated one or more times (especially with respect to x) to yield new series as convenient as the originals or nearly so. Examples are omitted. Usually when the correct form is found, even upon occasion by informal means, the recurrence relations (9) and (11) suffice to establish its general validity if it proves correct for two early consecutive members of the series.

Properties of P_n and P_n'

$$P_n(-u)=(-1)^nP_n(u) \tag{2}$$

$$P_n(\pm 1)=(\pm 1)^n \tag{3}$$

$$P_n(0)=0 \quad (n \text{ odd}) \tag{4}$$

$$P_n(0)=(-1)^{\frac{1}{2}n}\frac{1\cdot 3\cdot 5\cdot \cdots\ n-1}{2\cdot 4\cdot 6\cdot \cdots\ n} \quad (n \text{ even}) \tag{5}$$

$$P_n'(0)=nP_{n-1}(0) \tag{6}$$

$$P_n'(-u)=(-1)^{n-1}P_n'(u) \tag{7}$$

$$P_n'(\pm 1)=(\pm 1)^{n+1}\cdot\tfrac{1}{2}n(n+1) \tag{8}$$

$$(n+1)P_{n+1}+nP_{n-1}=(2n+1)uP_n \tag{9}$$

$$P_{n+1}'-P_{n-1}'=(2n+1)P_n \tag{10}$$

$$nP_{n+1}'+(n+1)P_{n-1}'=(2n+1)uP_n' \tag{11}$$

$$(1-u^2)P_n'=nP_{n-1}-unP_n \tag{12}$$

$$P_n\approx\left(\frac{2}{\pi n\sin\theta}\right)^{\frac{1}{2}}\left(1-\frac{1}{4n}\right)\sin[(n+\tfrac{1}{2})\theta+\pi/4] \tag{13}$$

$$\sin^2\theta P_n'\approx-\left(\frac{2n\sin\theta}{\pi}\right)^{\frac{1}{2}}\sin[(n+\tfrac{1}{2})\theta+3\pi/4] \tag{14}$$

Properties of Harmonics

$$(\partial/\partial x)(r^nP_n)=nr^{n-1}P_{n-1} \tag{15}$$

$$(\partial/\partial y)(r^nP_n)=-\sin\theta r^{n-1}P_{n-1}' \tag{16}$$

$$P_n=\frac{(-1)^n}{n!}r^{n+1}\frac{\partial^n}{\partial x^n}\left(\frac{1}{r}\right) \tag{17}$$

$$(1-u^2)P_n'=\frac{(-1)^{n-1}}{(n-1)!}r^n\frac{\partial^n u}{\partial x^n} \tag{18}$$

$$\partial/\partial u[(1-u^2)P_n']=-n(n+1)P_n \tag{19}$$

[8] See for example E. Jahnke and F. Emde, *Tables of Functions* (Dover Publications, New York, 1945), fourth edition.

Tables of Legendre Functions

Tallqvist,[9] Prévost,[10] and Mursi[11] have published the most useful tables[12] of P_n and of the associated functions P_n^m. Of the latter, the only functions of present interest are P_n^1 or in the notation here adopted $\sin\theta P_n'$. These functions, and therefore Mursi's table, will be most used. Note that $\sin\theta$ is tabulated as P_1^1. For n greater than 10, P_n^1 may be extrapolated from Mursi, using Eq. (11), or better Table III.§

Zeros of P_n and P_n'

These roots are significant in several connections. Those of P_n have been tabulated by E. R. Smith[13] to $n=40$, those of P_n' by Radau[14] to $n=10$. By empirical adjustment of Eqs. (13) and (14), extended series of both sets may be quickly calculated (Table II). This approximation checks P_n' zeros deduced from Tallqvist's table within ± 1.5 minutes of arc from $n=6$ to $n=32$ (± 0.0004 in u), with about half this error for P_n. Here θ_0 is the nearly constant angular interval between roots, which are θ_1, θ_2, $\cdots$, $\pi/2$, $\cdots$, $\pi-\theta_2$, $\pi-\theta_1$, counting from the axis; $\pi/2$ is omitted in the case of even P_n or odd P_n'.

Alternate Recurrence Formulas

Direct calculation of P_n or P_n' takes little longer than a second-order interpolation from tables. When all values of n are required, the standard forms (9) and (11) suffice, but usually only the alternate functions are

TABLE II. Approximate zeros of P_n and P_n' ($n>5$).

m	P_n $\theta_m-\theta_{m-1}$	$\theta_0=\pi/(n+.47)$ θ_m/θ_0	m	P_n' $\theta_m-\theta_{m-1}$	$\theta_0=\pi/(n+.51)$ θ_m/θ_0
1		1.218	1		0.766
2	1.012	2.230	2	0.992	1.758
3	1.004	3.234	3	0.998	2.756
4	1.001	4.235	4	0.999	3.755
>4	1.000	$m+.235$	>4	1.000	$m-1+.755$

[9] Hj. Tallqvist, Acta Soc. Sci. Fenn. (Helsingfors) Nov. Ser. A **2**, Nos. 4 (1937) and 11 (1938) $P_n(u)$, 6*D*, $n=1(1)16$, $u=0(0.001)1$. $P_n(\cos\theta)$, 6*D*, $n=1(1)32$, $\theta=0°(10')90°$. No differences.

[10] Georges Prévost, *Tables et Formules de Fonctions Spheriques* (Gauthier-Villars, Paris, 1933) $P_n(u)$, 5*D*, $n=1(1)10$, $u=0(0.01)1$. $P_n^1(u)$, 5*D*, $n=1(1)9$, $u=0(0.01)1$. First differences only. Reliable interpolation is difficult.

[11] Zaki Mursi, *Tables of Legendre Associated Functions* (E. and R. Schindler, Cairo, Egypt, 1941) (Stechert, New York). $P_n^1(u)$, 8*D*, $n=1(1)10$, $u=0(.001)1$. Modified second differences δ_m''.

[12] For other tables see Fletcher, Miller, and Rosenhead, *Index of Mathematical Tables* (McGraw-Hill Book Company, Inc., New York, 1946).

§ Again, if Tallqvist's table is available, this function may be found from it in any of three ways. Direct evaluation of the slope, using H. E. Salzer, *Table of Coefficients*, Nat. Bur. Stand. App. Math. Ser. **2** (1948) yields three decimals, but five may be obtained by use of Eq. (12). More than two applications of Eq. (10) to extrapolate an alternate series from Mursi may endanger the fourth decimal, because the errors accumulate.

[13] E. R. Smith, Iowa State Coll. J. Sci. **12**, 263 (1938), 6*D*.

[14] R. Radau, J. Math. **3** (6), 307 (1880), 8–10*D*.

TABLE III. Alternate recurrence coefficients.

	P_{n+2}		n Even			P_{n+2}		n Odd	
$P_0=1$	$P_2=\frac{1}{2}(3u^2-1)$				$P_1=u$	$P_3=\frac{1}{2}(5u^3-3u)$			
$n+2$	M	A	B	C	$n+2$	M	A	B	C
4	36	105	55	14	5	100	315	161	54
6	70	231	117	44	7	378	1287	649	260
8	616	2145	1079	450	9	312	1105	555	238
10	1350	4845	2431	1064	11	1870	6783	3401	1512
12	836	3059	1533	690	13	3276	12075	6049	2750
14	4186	15525	7775	3564	15	1750	6525	3267	1508
16	6480	24273	12151	5642	17	7888	29667	14849	6930
18	3162	11935	5973	2800	19	11286	42735	21385	10064
20	13300	50505	25271	11934					

	P_{n+2}'		n Odd			P_{n+2}'		n Even	
$P_1'=1$	$P_3'=\frac{1}{2}(5u^2-1)$				$P_2'=3u$	$P_4'=\frac{1}{2}(7u^2-3u)$			
$n+2$	M'	A'	B'	C'	$n+2$	M'	A'	B'	C'
5	20	105	49	36	6	140	693	333	220
7	90	429	209	130	8	462	2145	1053	630
9	728	3315	1635	952	10	360	1615	799	456
11	510	2261	1121	630	12	2090	9177	4557	2530
13	924	4025	2001	1100	14	1196	5175	2575	1404
15	4550	19575	9747	5278	16	1890	8091	4031	2170
17	2320	9889	4929	2640	18	8432	35805	17853	9520
19	3366	14245	7105	3774	20	3990	16835	8399	4446

needed. These can be found from Table III, which was developed by solving simultaneously three equations of the form of (9) or (11), using consecutive values of n. The result is

$$P_{n+2}(u)=\frac{1}{M}[(Au^2-B)P_n(u)-CP_{n-2}(u)]$$

with an analogous primed form for P_n'. The general coefficients are:

$$\begin{aligned}
M&=(n+1)(n+2)(2n-1)\\
M'&=n(n+1)(2n-1)\\
A&=(2n-1)(2n+1)(2n+3)\\
A'&=(2n-1)(2n+1)(2n+3)\\
B&=2n^2(2n+3)-1\\
B'&=2n(n+2)(2n-1)-3\\
C&=n(n-1)(2n+3)\\
C'&=n(n+1)(2n+3).
\end{aligned}\tag{20}$$

The table is complete through $n=20$, and since in each case the first two functions are given at the head of the column, it is also self-contained.||

|| In evaluating Legendre polynomials directly, the number of correct figures declines with rising order, for the terms alternate in sign and their individual magnitudes exceed the sum by larger factors as n increases. With the recurrence forms there is no systematic growth of legitimate rounding errors, but any lapse of the calculator is disastrous. Fortunately it is easy to devise simple over-all checks. For example, Table III is useful even when all values of n occur. Here the terminal values of each series, odd or even, involve every preceding member, while the two sets have no common basis. Thus a single application of (9) or (11) to the three terminal values is conclusive. Again, if the terminal value of P_n' lies within the range of Mursi's table, it may be checked directly, while beyond this range it is sufficient to apply (10) to the last two members of an alternate series, using Tallqvist.

THE AXIALLY SYMMETRIC CENTRAL FIELD AND ITS ERRORS

In this section the central field is examined apart from its origin. Such a disembodied approach separates the restrictions imposed by Laplace's equation from those of boundary conditions peculiar to the windings. When to suit experimental needs a possible type of central field has been chosen, in particular one in which the region of uniformity has a desired (and realizable) shape, there may remain much freedom of choice of the system geometry. This follows from the small number of possible variations in the field or its gradient, whether produced by isolated filaments or by massed windings, with or without equatorial gaps. When design proceeds outward from field specifications to the generating sources, it is less subject to irrelevant restrictions, and a more suitable shape or size of magnetic system may result.

The central field is defined as axially symmetric, bounded by a sphere that contains no sources, and with finite potential at the origin, which is the center of the sphere. Symmetry or antisymmetry about the equatorial plane is usually postulated; that is, each source is mirrored in the central plane, without or with reversal of current. This harmless restriction can be removed at any time by superposing a symmetrical and an antisymmetrical system.

The central potential, together with all its derivatives in cartesian coordinates, satisfies Laplace's equation and can be written as an infinite series of solid zonal harmonics. The subscript 0 will designate potentials or derivatives that are to be evaluated at the origin. Since V is finite, the central-field series contain only ascending powers of r. Without loss of generality one may set $V_0=0$, and so suppress the zero-order term. It is convenient to employ the exponent (m), in parentheses, to indicate m successive differentiations with respect to the axial coordinate x. From Eq. (15) it is seen that all the x-derivatives $V^{(m)}$ may also be expanded in ascending zonal harmonic series. But y as here used is not cartesian. Equation (16) shows that the series for y-derivatives contain derived polynomials like P_n' in place of P_n; the factor $\sin\theta$ satisfies the axial boundary condition $H_y=0$, imposed by the symmetry. On introducting the field constants h_n and the arbitrary factor $-1/n$, there results:

$$V=-\sum_{n=1}^{\infty}\frac{1}{n}h_n r^n P_n(\cos\theta).\tag{21}$$

This equation must conform, as a boundary condition, to the Taylor expansion of the axial potential V_{ax}, that is, to

$$V_{ax}=\sum_{n=1}^{\infty}\frac{1}{n!}V_0^{(n)}x^n.\tag{22}$$

Combine the last two equations with (3) and substitute

$H_z^{(0)}$ for $-V^{(1)}$:

$$H_{z,0}^{(n-1)} = (n-1)!h_n. \tag{23}$$

Thus each of the field constants h_n is proportional to one x-derivative of the axial field at the origin. Evidently h_n receives contributions from all the field sources, and this dependence is conveniently expressed in terms of a set of dimensionless source constants q_n (Eqs. (43) to (54)) defined from the relation

$$q_n \equiv cr_0^n h_n/2\pi i \quad \text{(gaussian units)}. \tag{24}$$

As previously noted, r_0 is an arbitrary reference radius. Since at this stage the sources must not impose boundary conditions upon the field, further discussion of the source constants is deferred.

The fields H_r and H_θ may be written down at once by differentiating (21); H_x, H_y, and their higher derivatives with respect to cylindrical coordinates x, y are found from the same equation by applying (15) and (16).

$$H_z^{(0)} = \sum_{n=1}^{\infty} h_n r^{n-1} P_{n-1}(\cos\theta) \tag{25}$$

$$H_z^{(1)} = \sum_{n=2}^{\infty} (n-1) h_n r^{n-2} P_{n-2}(\cos\theta) \tag{26}$$

$$H_z^{(m-1)} = \sum_{n=m}^{\infty} \frac{(n-1)!}{(n-m)!} h_n r^{n-m} P_{n-m}(\cos\theta) \tag{27}$$

$$H_y = -\sum_{n=2}^{\infty} \frac{1}{n} h_n r^{n-1} \sin\theta P_{n-1}'(\cos\theta) \tag{28}$$

$$\partial H_x/\partial y = \partial H_y/\partial x = -\sum_{n=3}^{\infty} h_n r^{n-2} \sin\theta P_{n-2}'(\cos\theta). \tag{29}$$

Equation (27) may be written

$$H_z^{(m-1)} = (m-1)!\, h_m + \cdots. \tag{30}$$

The series receives no contribution from any term of (21) before the mth. Its first term agrees at the origin with (23), but it offers a broader interpretation of the field constants. Each measures the strength of a particular x-derivative, not merely at the origin, but throughout the region since no field coordinates appear in the first term of Eq. (30). Its later terms are dismissed as belonging to field constants of higher order than the mth. Thus any axially symmetric central field is fully described by stating the amplitudes h_n of a set of standard harmonic components. Harmonic h_1 is a uniform field; h_2 is a uniform gradient $H_z^{(1)}$, with H_y proportional to $r\sin 2\theta$, and so on. There is a certain resemblance to fourier analysis, but here each component is a three-dimensional field pattern, while the amplitudes depend upon the choice of origin. Figures 2, 3, and 4 show contours of constant $H_z^{(0)}$ in fields of types h_3, h_5, and h_7, respectively, that is in fields for which $H_z^{(2)}$, $H_z^{(4)}$, or $H_z^{(6)}$ is constant. More generally, they are contour maps of $H_z^{(m)}$ in fields h_{m+3}, h_{m+5}, h_{m+7}, where m may be odd or even. Harmonic types with odd h_n are characterized by symmetrical H_z patterns, with antisymmetrical x-gradient and H_y patterns, and conversely.

No direct use has been suggested for field types of order higher than the gradient h_2, though it is easy to produce a field that reduces to h_3 or h_4 near the origin, and in theory any harmonic is easily generated in pure form. To produce h_n, wind a sphere with surface current density proportional to $\sin\alpha P_n'(u)$, that is with axial current density proportional to P_n'. The single component $H_z^{(n-1)}$ is then uniform throughout the sphere. A later section discusses systems of circular filaments that approximate such current sheets.

Field or Gradient Errors and Error Maps

The error η_e of order e is arbitrarily defined, for field systems, as the dimensionless ratio of the error-field $H_z^{(0)}$ contributed by the term of order $e+1$ to the fundamental field $H_{z,0}^{(0)}$ due to h_1, while the corresponding ratio for gradient systems is that of the components $H_z^{(1)}$ due to h_{e+2} and to h_2. In field analysis, it is convenient to write the errors in terms of coefficients k_e derived from the field constants, while for coil design a more appropriate set of error coefficients ϵ_e may be defined from the source constants. The latter set, like q_n, are pure numbers. From Eqs. (24)–(26),

$$\eta_e = k_e r^e P_e(\cos\theta) = \epsilon_e (r/r_0)^e P_e(\cos\theta) \tag{31}$$

where

$$k_e = h_{e+1}/h_1 \qquad \epsilon_e = q_{e+1}/q_1 \qquad \text{(Field)}$$

$$k_e = (e+1)h_{e+2}/h_2 \quad \epsilon_e = (e+1)q_{e+2}/q_2 \quad \text{(Gradient)}. \tag{32}$$

Assuming a plane of symmetry or its converse, $e = 2, 4, 6, \cdots$ and the errors are solid zonal harmonics of even order. The equal error contours conform to the equation

$$r^e P_e(\cos\theta) = \text{const} \quad (e = 2, 4, 6, \cdots). \tag{33}$$

Therefore, all contours of a given order are geometrically similar to one another, and are the same for field as for gradient systems; they are identical for all systems, regardless of the number and complexity of their sources. In fact they are identical with certain electrostatic equipotentials and with a number of other surfaces in various Laplacian fields, over which some scalar magnitude is constant.

The order of error of any system is defined to be that of its lowest order error, measuring the gap, if any, in its even or odd harmonic expansion; if m terms are missing, the order of error is $2m+2$. Universal error maps may be drawn to cover all cases in which interest is confined to a region so near the origin that the lowest error term alone is significant, or alternatively the first two. Figures 2, 3, and 4 cover single-error fields of the

FIG. 2. Second-order error contours for central field or gradient.

second, fourth, and sixth orders, respectively,¶ while Figs. 5 and 6 represent two-error hybrid fields. The contour interval is logarithmic. In the first three figures the errors increase by powers of ten (for example, 0.01 percent, 0.1 percent, 1 percent), and in the last two by powers of two. Where three levels are shown, the innermost set is merely indicated by short arcs. The most elementary magnetic fields to which these diagrams correspond are (Fig. 2) Ampere's loop, (Fig. 3) the Helmholtz field pair or Maxwell's gradient pair, and (Fig. 4) Maxwell's three-loop system.[15]

The H_y error, defined for field systems as $H_y/H_{z,0}$, may be similarly mapped. The contours are given by

$$r^e \sin\theta P_e'(\cos\theta) = \text{const} \quad (e=2, 4, 6, \cdots) \qquad (34)$$

and they show a general resemblance to those for H_z, but with the axes rotated through an angle approximately $\pi/2e$. In this case the axis and the mid-plane are limiting cases of zero-error cones, instead of representing error maxima.** Equations (13), (14) with (25) and (28) show that for all values of e the magnitude of the error H_x is much the same as that of H_y.††

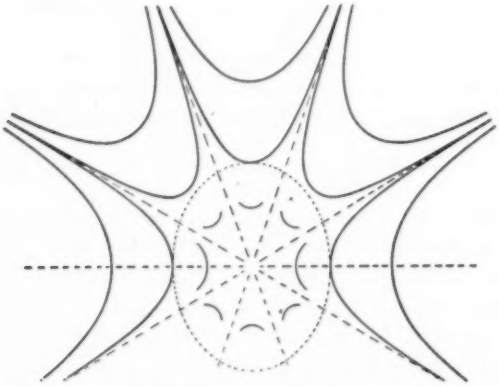


FIG. 3. Fourth-order error contours with error spheroid.

In each figure, the zero-error contour for H_z or $H_z^{(0)}$ is a set of cones with e sheets intersecting at the origin, with half-angles corresponding to the roots of $P_e=0$ (not shown in Fig. 4). The cones separate compartments whose errors alternate in sign, but no signs are shown since each field occurs in positive and negative variants. McKeehan[16] has discussed the cones, but not the remaining contours. Principal interest attaches to the shape of the low error region. For stated limits such as 1 percent, the available space is best described as cylindrical in second-order fields, while it is more nearly spheroidal for all higher orders. In fact, a true spheroid, tangent to the axial and equatorial error sheets, always clears safely the intervening oblique sheets of a single-error field. Such a tangent ellipsoid (as shown in Fig. 3) is called hereafter the error spheroid, and is considered to bound the useful region. When as in Figs. 5 and 6 the contours are warped by superposed higher orders,

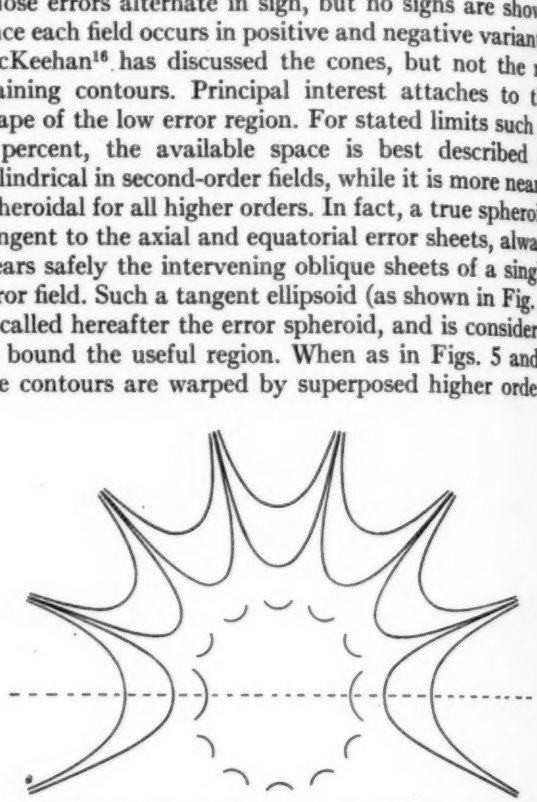


FIG. 4. Sixth-order error contours, omitting cones.

they generally lose the spheroidal form, and each assigned error limit bounds a differently shaped space.

The aspect ratio of the error spheroid of order e, defined as the ratio of the axial to the equatorial diameter, is $[P_e(0)]^{1/e}$. As this is less than unity, all single-error central fields are oblate. Table IV lists the aspect ratios from the second to the twelfth order. The plus sign which is written when e is a multiple of 4 means that the sign is the same for the axial and equatorial errors (because the error reverses sign at each of $\frac{1}{2}e$ zero-error cones). To illustrate the comparative ad-

¶ From the trend of the figures, the higher orders are easily visualized. Compare Fig. 3 of Shortley and May (reference 3) which shows the 8th-order two-dimensional contours of the cylindrical harmonic $r^e \cos e\theta$ ($e=8$). For the simplest electrostatic equipotentials of second order (Fig. 2, present paper) place equal point charges $+Q_1$ at $\pm x_1$, 0. For Fig. 3, add a pair $-Q_2$ at $\pm x_2$, 0, where $Q_1 x_2^3 = Q_2 x_1^3$. The analogy may be improved and extended by substituting charged rings for the axial point charges.

[15] J. C. Maxwell, *Treatise on Electricity and Magnetism*, Vol. II, Sec. 715.

** In the two-dimensional case (Fig. 3 of reference 3) the H_y errors are proportional to the cylindrical harmonic $r^e \sin e\theta$, and thus are identical with the H_z errors, as mapped, rotated through an angle $\pi/2e$.

†† H. Nagaoka, Phil. Mag. **41**, 377 (1921, plates follow p. 548) drew second- and fourth-order maps for the Ampere and Helmholtz fields, separately, both H_z and H_y, but his sixth-order map is meaningless. Here a numerical error in the complicated derivation from elliptic integrals changed one coefficient from 35 to 1235! Nagaoka did not notice that his results finally emerged as zonal harmonics expressed in x, y, nor did he ascribe to these maps of special cases a more general validity.

[16] L. W. McKeehan, Rev. Sci. Instr. **7**, 150 (1936).

vantage of high order fields for extreme uniformity, note that if in Fig. 2, as in Fig. 4, an inner third contour were shown, it would embrace only one percent of the area of the outer contour. On the other hand, high order fields have increasingly sharp boundaries, as the errors involve higher powers of r. This is shown in Table IV, where the term "limit ratio 10:1" means the ratio of diameters of two error spheroids having errors, respectively, 10η and η. Such ratios are even roots of 10. Table IV when added to the error maps gives an adequate summary of all single-error fields.

Can all fields be cast in so rigid a mold? There are three reservations. First is the limitation to a spherical space that contains no field source. Though (subject to reservation two) only the innermost end-turn of a smooth solenoid bounds the field, the nearest of a series of openly spaced turns must be taken. However, if the spacing is uniform and close, the system may be considered smooth and the gaps superposed as a perturbation, as is often done to estimate the effect of winding errors.

Second is the boundary condition encountered at a current sheet, such as the wall of a thin solenoid. Here fictitious fields H_x' and H_y' are first calculated to the previously defined limits of the central region. If N' is

TABLE IV. Aspect and limit ratios.

Order of error	2	4	6	8	10	12
Aspect ratio	−0.707	+0.783	−0.824	+0.850	−0.869	+0.883
Limit ratio 10:1	3.162	1.778	1.468	1.334	1.259	1.212

the linear turn density, the true fields are then given by

$$H_y=H_y' \quad H_x=H_x'-4\pi N'i/c. \tag{35}$$

For, at all points in the field of an arbitrary cylindrical current sheet with normal end-faces (not restricted to circular cylinders) the axial component is

$$H_z=\Omega N'i/c \tag{36}$$

where Ω is the solid angle subtended by the current sheet, regarded as a gaussian surface but with open ends. This angle varies continuously except for the sudden increment 4π at the wall. Thick solenoids require stepwise increments, layer by layer, as the wound space is traversed. With these modifications, the error contours continue through the walls as if they were absent, but each contour represents two distinct levels spliced together at the wall, differing from each other by $4\pi N'i/c$.

Hybrid Fields, Prolate and Oblate

The third reservation is the restriction to a single order of error. Removal of this restriction permits minor alterations in the aspect ratio of the useful field, for some specific error tolerance that is preassigned. Variations resulting from combination of two or more error terms are more limited than one might expect, for the

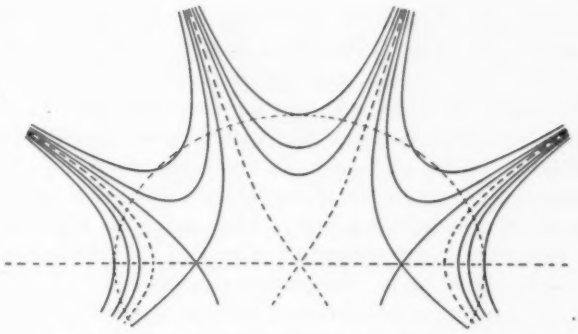


FIG. 5. Combined second- and fourth-order errors, with error spheroid; prolate case.

significant features all appear in the prolate and oblate field maps of Figs. 5 and 6. The writer knows of no proposed hybrid in which the second order is absent, though the success of the oblate field of Fig. 6, mixing second and fourth, suggests that a proper combination of fourth- and sixth-order errors without the second might be even better. Present discussion concerns the second and fourth orders alone.‡‡ Though many have employed the prolate field of Fig. 5, especially when using solenoids with over-correction of the end errors, the oblate field is more useful. It was introduced by Shaw,[17,18] and the minima in Koch's curves[7] for thick cloud-chamber coils correspond to the same field type, though Koch's data do not readily yield any over-all field pattern. Either prolate or oblate fields may be generated by a variety of systems; some facts about each type have been given, but more systematic analysis is lacking. The following treatment applies equally to systems producing either fields or gradients. With these hybrids added to the single-error types, it may be doubted that further basic central fields exist.

If two consecutive orders have the same sign of error

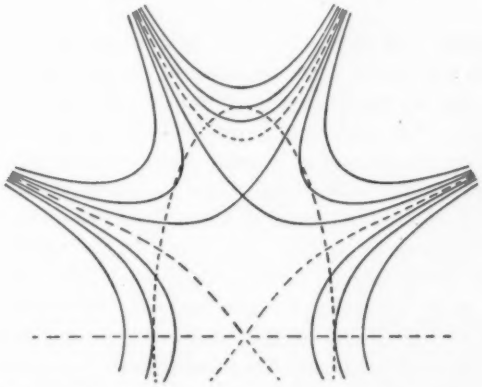


FIG. 6. Combined second- and fourth-order errors, with error spheroid; oblate case.

‡‡ In a later paper, a field similar to that of Fig. 5, produced by a pair of thick-solenoid cloud-chamber coils, will be calculated through the tenth order. Little change results from this extension, because as shown below (Eqs. (55)–(58)) the initial terms converge rapidly.

[17] A. E. Shaw, Phys. Rev. **54**, 193 (1938).

[18] Ference, Shaw, and Stephenson, Rev. Sci. Instr. **11**, 57 (1940).

on the axis they differ conversely at the equator. The lower order error is dominant near the origin, but the higher order varies as a higher power of r and soon outweighs it. Thus with one sign of the higher order, its addition accelerates the normal monotonic growth of error on the axis, while in the plane $x=0$ the error reaches a maximum, declines to zero, changes sign, and thereafter increases rapidly. The region of low error is extended equatorially but is axially compressed, becoming more oblate. The prolate case requires a change of sign of one error pattern; since it is similar except that the axis and the equatorial plane exchange roles, the two cases may be discussed together.

Let $0, y_m$ define the equatorial circle, or $\pm x_m, 0$ the axial points, at which the error reaches its maximum value, while at $0, y_0$ or at $\pm x_0, 0$ it has fallen again to zero. For the second and fourth orders, Eq. (31) reduces to

$$\eta = k_2 r^2 P_2 + k_4 r^4 P_4. \tag{37}$$

From this the maxima and the zeros are found to be as follows:

$$y_0^2 = 4k_2/3k_4, \quad x_0^2 = -k_2/k_4,$$
$$y_m^2 = 2k_2/3k_4, \quad x_m^2 = -k_2/2k_4 \tag{38}$$

TABLE V. Comparative dimensions of error spheroids.

	Prolate Fig. 5	Spherical Fig. 5	Normal Fig. 3	Oblate Fig. 6
Aspect ratio	1.22	1.00	0.78	0.38
Equatorial limit	0.82	0.89	1.00	1.39
Axial limit	1.27	1.14	1.00	0.67
Useful volume	0.86	0.92	1.00	1.30

while the magnitudes of the maximum errors are

$$\eta_{m,y} = -k_2^2/6k_4, \quad \eta_{m,x} = -k_2^2/4k_4. \tag{39}$$

Equation (38) shows that k_2 and k_4 must be of the same sign in the oblate case, opposite in the prolate. In either case the zero-error contour (dashed in the figures) is divided. The first part is asymptotic at the origin to the cone $P_2=0$ and at remote points to one of the two P_4 cones. The remaining parts, through y_0 or x_0, are asymptotic at far points to the second cone. Thus each drawing reduces to Fig. 2 near the origin and to Fig. 3 at the periphery. The drawings are again universal, their scale being set by the ratio k_2/k_4, which determines the positions of the zeros and the maxima. The interval is logarithmic, the error doubling at each step. Thus the contours drawn are $\pm\eta_m$, $\pm 2\eta_m$, $\pm 4\eta_m$, of which $+\eta_m$ intersects itself in the circle y_m or at the points $\pm x_m$. The entire region between the dashed contours has the same sign of error as the branched contour $+\eta_m$ that lies within it, while errors of opposite sign are encountered everywhere else. To construct the maps, the quadratic (37) was solved for a considerable number of angles.

Maximum field uniformity occurs wherever the contours branch. Contours near the singular points $\pm x_m$ or near the origin have the exact form shown in Fig. 2. The circle of radius y_m is an ideal orbit for charged particles, and it was so used by Shaw. In this circle two cones intersect at right angles. Since the error is maximum, $\partial H_z/\partial y = 0$, while the remaining first derivatives vanish from symmetry. Second-order variations alone remain and the circle is the axis of an annular region of excellent uniformity. The annulus has, for a given error, approximately the same width radially and axially, as may be proved directly from Laplace's equation in cylindrical coordinates. Since $\partial H_z/\partial_y = 0$, the second derivatives of H_z with respect to x and y are equal and opposite. Of these $H_z^{(2)}$ can be found from Eq. (27), retaining only the terms in $h_3 = k_2 h_1$. Substitute $P_2 = -\frac{1}{2}$ and insert the value of $r^2 = y_m^2$ from (38); the result is

$$\partial^2 H_z/\partial y^2 = \partial^2 H_z/\partial x^2 = 2k_2 h_1 = 2h_3. \tag{40}$$

The foregoing theory is easily extended to include higher orders.

In these hybrid fields, the shape of the useful space depends on the magnitude of the tolerable error. Assume that errors of either sign are allowed, up to a stated maximum η_0, and that a spheroidal working space is desired. Previous writers have identified η_0 with η_m, but the figures show that the oblique sheets pinch off part of the space enclosed by $\pm\eta_m$. To be quite safe, η_m must be reduced to $\frac{1}{2}\eta_0$ in oblate fields, and to $\frac{1}{4}\eta_0$ in prolate fields. Comparison of the normal fourth-order spheroid of Fig. 3 with those of Figs. 5 and 6, which correspond to the values of η_0 just named, shows that the presence of a second-order error may change the aspect ratio significantly.

The design procedure is, starting with Fig. 3 ($k_2=0$) to modify the aspect ratio by slight adjustment of position or dimensions of some part of the source system, whereby the proper small value of k_2 is realized. In this process, which requires Eqs. (43)–(54), (24) and (32), k_4 changes little; both η_0 and k_4 will be considered constant. To compute Table V, set $4\eta_0 = \eta_{m,x}$ or $2\eta_0 = \eta_{m,y}$, and solve for axial and equatorial values of r. These are the semi-axes of the spheroids of the figures. All quantities except the aspect ratio express ratios of the modified dimensions to those of the normal fourth-order case, for the same tolerable error η_0. The spherical case is derived from Fig. 5 by selecting a set of contours beyond η_m in which a sphere may be inscribed. The sphere passes well inside the oblique sheets. Though for special purposes these hybrid fields are superior to pure fourth-order fields, the sixth-order solenoid forms described at the end of this paper are still better, and may prove equally easy to construct.

CALCULATION OF THE SOURCE CONSTANTS

The central field of any axially symmetric system is fully determined when the source constants of Eq. (24)

can be computed. The unit source is a filament f_j, whose field, at the origin, is

$$V_0=-iN_j\Omega/c=-2\pi iN_j(1-u_j)/c. \quad (41)$$

Expand this equation along the axis, using Eq. (22), without the arbitrary constant V_0. Then eliminate $u_j^{(n)}$ with the aid of Eq. (18). Here drop the factor $(-1)^{n-1}$, since the x-differentiation in (18) implies axial motion of the source, whereas in the Taylor expansion (22) it refers to motion of the field point. The result is

$$V_{ax}=-\frac{2\pi i}{c}N_j\sin^2\alpha_j\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{x}{r_j}\right)^n P_n'(u_j). \quad (42)$$

This must be fitted as a boundary condition to the harmonic expansion (21), using also Eq. (3). On writing the result in terms of the constants defined by Eq. (24), at the same time summing over u_j, the source constants $q_n(f)$ are found, for a complete filament system:§§

$$q_n(f)=\sum_j N_j\sin^2\alpha_j(r_0/r_j)^nP_n'(u_j). \quad (43)$$

To deduce $q_n(s)$ for a solenoid system, rewrite the summation as an integral, replacing the element N_j by $N_k'dx$. After q_1, the integrals are found from (18). Insertion of the limits introduces the coordinates of the end-turns alone of the thin-solenoid system.[19]

$$q_1(s)=r_0\sum_k N_k'u_k$$

$$(n>0),\ q_{n+1}(s)=\frac{-r_0}{n}\sum_k N_k'\sin^2\alpha_k(r_0/r_k)^nP_n'(u_k). \quad (44)$$

The following equation relates the source constants for one end of a solenoid of strength N_s' and those of a coincident filament N_f:

$$(n>0),\quad N_fq_{n+1}(s)=-(r_0/n)N_s'q_n(f). \quad (45)$$

This follows from the last two equations, or from (36) and (41), which express $V(f)$ and $H_x(s)$ in terms of the same solid angle. More generally, x-differentiation transforms a cylindrical current sheet into a circular filament, a thick solenoid into a spiral disk (Eq. (47)), or any of the quantities h_n, q_n, r^nP_n or $r^{-n}\sin^2\alpha P_n'$ into the similar quantity of next higher order. Especially useful is the series relation among the constants q_n. From Eqs. (18) and (43),

$$q_{n+1}=-(r_0/n)\partial q_n/\partial x. \quad (46)$$

Here q needs no specific argument, as (f) or (c), since the equation involves no source coordinates. It holds for each circular turn, hence also for every type of complete axially symmetric system.

In a later paper, Eqs. (43), (44) will be expanded as power series in the section coordinates of moderately thick solenoids, referred to origins at the central filament of each square section or at the central filament of each end-face of a more general section. Such series for q_n are sometimes useful, but the same constants may also be found by integrating Eqs. (43), (44) with respect to a, yielding (48) which is valid for sections of unrestricted thickness. From (43) are derived source constants $q_n(d)$ for a system of plane annular current sheets (transverse disks), while from (44) come the thick-solenoid forms. Since the disk is a thin slice from the end of a solenoid, a relation like (45) holds for the two sets of coefficients:

$$(n>0),\quad N_d'q_{n+1}(c)=-(r_0/n)N_c''q_n(d) \quad (47)$$

where N_d' and N_c'' are, respectively, turns per (radial) cm and turns per cm², and the disk edges coincide with those of the solenoid. It is easy to approximate the disk by rolling up a narrow tape, in the manner described by Bitter[20] and others,[21] but the possible advantages of such a geometry are unknown.

TABLE VI. Coefficients for series s_{n+2}.

$n+2$	A_0	A_2	A_4	A_6	A_8	A_{10}	A_{12}
3	1						
4	1	3					
5	2	3	15				
6	2	3	0	35			
7	8	12	15	−70	315		
8	8	12	15	35	−315	693	
9	16	24	30	35	315	−2079	3003

The limits of integration introduce coordinates of the circular edges; the outer right-hand edge of a positive winding takes the plus sign. From Eqs. (44), (46), $q_n(c)$ is easily found. Only $q_1(s)$ is directly integrated, while the higher orders require only successive differentiations. The source constants will first be written in powers of x rather than of r_l, because in practice it is convenient to add contributions from the two edges of each end-plane before multiplying by x_l^n.

$$q_1(c)=r_0\sum_l N_l''x_l\ln\left(\frac{1+\sin\alpha_l}{u_l}\right)$$

$$q_2(c)=r_0^2\sum_l N_l''\left[\sin\alpha_l-\ln\left(\frac{1+\sin\alpha_l}{u_l}\right)\right]$$

$$(n>0),\quad q_{n+2}(c)=\frac{r_0^2}{n(n+1)}\sum_l N_l''(r_0/x_l)^ns_{n+2}(u_l). \quad (48)$$

§§ These differ only by a constant from McKeehan's S_n (reference 16).

[19] The method, though not the form, is essentially that of Andrew Gray, Phil. Mag. (V) 33, 62 (1892). See also Andrew Gray, *Absolute Measurements in Electricity and Magnetism* (MacMillan Company, Ltd., London, 1921) 2nd ed., pp. 201–206, 796–797. A certain lack of organization in Gray's treatment may explain the general neglect of this simple method. It was used by Dwight (reference 6) to derive a number of formulas with restricted choice of origin.

[20] F. Bitter, Rev. Sci. Instr. 7, 482 (1936).

[21] L. G. Coonrod *et al.*, Rev. Sci. Instr. 20, 408 (1949).

TABLE VII. Coefficients of $2048U_{n+2}$.

$n+2$	u	u^3	u^5	u^7	u^9	u^{11}	u^{13}	u^{15}
3	+3072	−768	−128	−48	−24	−14	−9	−6.
5	−11520	+21760	−6000	−1056	−410	−210	−125.	−81.
7	+22400	−1 09200	+1 27344	−29260	−4695	−1726.	−853.	−494.
9	−35280	+3 24576	−7 99092	+6 73668	−1 28555.	−18321.	−6207.	−2892.
11	+49896	−7 41510	+30 84345	−50 88817.	+33 56557.	−5 42992.	−68764.	−21333.
13	−66066	+14 44014	−90 36590.	+239 32015.	−297 02787.	+160 82329.	−22 50561.	−2 55127.

Here $s_{n+2}(u)$ is a summation that is usually best computed as follows:

$$s_{n+2}(u)=1-\sin^3\alpha\sum_{m=1}^{m=n}u^{m-1}P_m'(u)\equiv 1-\sum_{m=1}^{m=n}f_m(u). \quad (49)$$

The second half of (49) defines a series $f_m(u)$ from which $s_{n+2}(u)$ is computed as a running sum. The first half expresses compactly the general series that results from repeated differentiation of q_2 in Eq. (48), though the early terms are more easily derived as polynomials (Eq. (50)) whose coefficients are listed in Table VI.

$$s_3(u)=1-\sin^3\alpha \quad s_4(u)=1-\sin^3\alpha(1+3u^2)$$

$$s_{n+2}(u)=1-\frac{\sin^3\alpha}{A_0}\times(A_0+A_2u^2+A_4u^4+\cdots+A_{2n-2}u^{2n-2}). \quad (50)$$

The series f_m of Eq. (49) may be extended indefinitely by using a recurrence formula based on Eq. (11):

$$f_1=\sin^3\alpha \quad f_2=3u^2\sin^3\alpha$$

$$f_m=\frac{u^2}{m-1}[(2m-1)f_{m-1}-mf_{m-2}]. \quad (51)$$

When m is 10 or less, the last function, $f_m=\sin^3\alpha u^{m-1}P_m'(u)$, may be checked from Mursi's table. This verifies the entire series, and suffices for five errors of the system, whether for field or gradient. The function $u^{-n}s_{n+2}(u)$ is roughly of the order of magnitude n^4, for all values of u. Let r_0/r_1 be of the order unity. When u is small, r_0/x_1 is large, s_{n+2} diminishes rapidly with increasing n, but the decrease is offset by the rapidly rising factors $(r_0/x_1)^n$. The convenience of Eq. (48) is lost, and it is better to rewrite it in closer analogy to (44), even though each source now requires a separate power series in r_1.

$$(n>0), \quad q_{n+2}(c)=\frac{r_0^2}{n(n+1)}\sum_1 N_1''(r_0/r_1)^nU_{n+2}(u_1) \quad (52)$$

where

$$U_{n+2}(u)\equiv u^{-n}s_{n+2}(u)=u^{-n}-\sin^3\alpha\sum_{m=1}^{m=n}u^{m-n-1}P_m'(u). \quad (53)$$

The function U_{n+2} is computed as follows:

$$U_3(u)=u^{-1}(1-\sin^3\alpha)$$

$$U_n(u)=u^{-1}U_{n-1}(u)-\sin^3\alpha P_{n-2}'(u). \quad (54)$$

The series $-\sin^3\alpha P_n'$ is developed from $-\sin^3\alpha$, using Eq. (11).

In computing, either from Eqs. (52)–(54) or from (48)–(51), if u is too small the number of significant figures declines with advancing n, since the absolute errors accumulate in a running total of decreasing magnitude. The remedy is to multiply Eq. (50) out in full, after expanding $\sin^3\alpha$ by the binomial theorem. Table VII lists enough coefficients of the resulting infinite power series in u to satisfy all reasonable demands for precision, up to q_{13}, for u as high as 0.5. This formidable table is rarely needed in practice for u greater than perhaps 0.25, and then only a single line of it, from which some coefficients at the right end may usually be omitted. The best method is to enter the table, when necessary, at the last value of $n+2$ that is required, e.g., at 11 if five error terms are needed. Having computed the terminal coefficient to about four *digits*, proceed backward using (54) or (49) and retaining a constant number of *decimals*. Check the terminal value s_3 or U_3 by computing $\sin^3\alpha$ directly. This check confirms all results, and all will be found to have enough significant digits. Table VII gives coefficients of u_m in the series $2048U_{n+2}$; to find s_{n+2}, multiply by $u^n/2048$. The factor 2^{11} allows a reasonable number of integral coefficients. Only those at the lower right, with decimal point appended, have been rounded off.

The foregoing treatment of thick solenoids has some relation to that of L. W. McKeehan, Rev. Sci. Instr. **10**, 371 (1939). Here occur two functions equivalent to $q_3(c)$ and $q_5(c)$. The actual expressions were so complicated that the series was little used, and no attempt was made to extend it further. Actually the three-term equivalent of s_3 can be reduced by trigonometric substitutions to $\sin^3\alpha$, and presumably the nine-term function would similarly reduce to s_5.

H. W. Koch (reference 7; Koch cites also previous work) has recently written out the first four error terms of two rather difficult series for thick solenoids. Together, they cover the range $\alpha>\pi/4$, giving fields at points in the plane of the origin. From these series he computed a table, as a basis for further graphical analysis. The table gives H_x at equatorial points for solid-wound cylinders, corresponding to single sources c_1, but Koch's expressions do not reduce readily to a form like Eq. (48), from which presumably they differ by some function of x_1. This is inferred from numerical comparisons that included over a dozen selected tabular entries. Only when two tabular values with the same x-coordinate were added algebraically to yield H_x for a complete solenoid end-plane did the results agree with calculations from Eq. (48). With this restriction, the last figure checked within one or two units, but second and fourth differences formed from the complete table reveal at least one large error, with one or two small ones. From various comparisons it was concluded that the rate of convergence of Koch's series is much the same as that of the more versatile and convenient series developed herein.

At all stages of this work, numerical checks were liberally employed. Errors were computed at the same point, using different origins. Calculations from thick-solenoid formulas were checked against gaussian numerical integrations using four or six thin solenoids of suitable strength properly distributed through the section, and against fourth-order section expansions of median thin solenoids or filaments. Limiting cases were analyzed, and many other special tests were devised.

RATE OF CONVERGENCE OF ZONAL HARMONIC SERIES

The discussion here concerns the rate of convergence of the complete series giving, for example, H_z or H_y at a central or remote field point, or the mutual inductance M of two systems. Assuming the usual alternate series, the rate is measured by the ratio of the contribution from order $n+2$ of such a series to that from order n. Complete series that involve both field and source coordinates are not written out explicitly in this paper, but they may be derived from Eqs. (25)–(29) by eliminating h_n, using Eqs. (24) and (43), (44), or (48). Analogous series occur for remote fields and for mutual inductances. It seems at first sight difficult to treat the general case concisely, since each of these three types of series has distinct forms for filament systems, for thin and for thick solenoids. But all conform to a common pattern, in that each term is the product of factors of three kinds, each with its own habit of convergence. Over-all constants are dismissed as irrelevant.

The first factor, which of itself converges steadily, is a power of a ratio of two polar radii; for example, the ratio of the field-point radius to a source-point radius, as r/r_j when combining Eqs. (25)–(29) with (43). In the corresponding remote-field case the ratio is inverted. Though the actual factors q_n, p_n involve summations over all sources, the rate of convergence after a few terms depends mainly upon the more persistent contributions from the nearest source in the case of q, or from the most remote in the case of p. Mutual inductance series contain products of q_{n1}, from the outer or primary system, and p_{n2}, from the secondary inner system. The persistent terms arise from source pairs of minimum r_1 and maximum r_2. It is therefore a safe approximation to discard all sources except those least favorably placed, and from these to define the r-dependent factor ρ as the ratio r/r_j, r_j/r, r_2/r_1, etc.

The second factor is a product of reciprocals of integers nearly equal to the term order n, as n, $n\pm1$, $n\pm2$. In considering the ratios of consecutive terms, the effect of the increment ±1 or ±2 is of second order, decreasing rapidly with rising n. Write, therefore, m for any of these integers, and m^p for any of their products of degree p.

The third factor is a product of two (sometimes three) angle-dependent functions. Dropping the distinction between θ and α, these may be P_m, $\sin\theta P_m'$, $\sin^2\theta P_m'$, U_m, or a fifth function that replaces U_m in the remote field. All exhibit oscillations-in-m of quasi-sinusoidal character. The asymptotic behavior of P_m and of $\sin^2\theta P_m'$ is described by Eqs. (13), (14); that of U_m and its analog is not known with precision, but these functions seem to behave like $\sin^2\theta P_m'$, and such a property will be tentatively assumed here. When m varies, P_m and $\sin^2\theta P_m'$ oscillate at nearly the same rate as $\sin m\theta$ or $\cos m\theta$, i.e., with period-in-m close to $2\pi/\theta$, but with amplitudes nearly proportional to $m^{-\frac{1}{2}}$ and to $m^{\frac{1}{2}}$, respectively.‖ ‖ The two functions are related like sine and cosine in that when θ varies the zeros of each coincide with maxima of the other (Eq. (19)) while when m varies the two functions are in quadrature (Eqs. (13), (14)). Write $S(m\theta)$ for a sinusoidal function of constant amplitude and undefined phase, with period like that of $\sin m\theta$, i.e., $2\pi/\theta$. Interpret the symbol $\approx$ to mean "converges as," that is let it be used to connect quantities that depend similarly upon m and that converge at similar rates. Then,

$$m^{\frac{1}{2}}P_m\approx m^{-\frac{1}{2}}P_m'\approx m^{-\frac{1}{2}}U_m\approx S(m\theta). \tag{55}$$

It is expedient to replace q_m or p_m by a generalized source constant c_m. Equations (25)–(29) now show that H_z, H_y, with their gradients and all higher derivatives, converge at similar rates. Writing H for H_z or H_y,

$$H_n^{(m)}\approx m^{-\frac{1}{2}}c_m\rho^m S(m\theta). \tag{56}$$

From Eqs. (43), (44), (48) and from their remote-field analogs,

$$c_n(f)\approx m^{\frac{1}{2}}S(m\theta),\quad c_n(s)\approx c_n(d)\approx m^{-\frac{1}{2}}S(m\theta),$$
$$c_n(c)\approx m^{-\frac{1}{2}}S(m\theta). \tag{57}$$

The series for mutual inductance (from a later paper) contain $m^{-2}q_{n1}p_{n2}$, and, when the axes of the primary and secondary systems intersect at angle θ, $P_n(\cos\theta)$ also. Since the last factor may be absent, it will be enclosed in brackets, thus:

$$M_n\approx m^{-5/2}\rho^m c_n^2[S(m\theta)]. \tag{58}$$

The rate of convergence for any complete series can be estimated by substituting (57) into (56) or (58). For a given field point, the resulting series contain products $S(m\theta_1)S(m\theta_2)$, with periods respectively, $2\pi/\theta_1$, $2\pi/\theta_2$; such products oscillate with "beats." The oscillations are "damped" only by the factors in ρ and in m, so that any monotonic decrease must depend upon these factors rather than upon the choice of angles or combinations of angles. This fact has sometimes been misunderstood.

The ratio of the "amplitudes" alone of successive all-odd or all-even terms is $\rho^2(m/m+2)^p$, where p ranges from zero for the field produced by filaments to 11/2 for the mutual inductance of two thick solenoids. The m-factor may thus influence strongly the convergence of the first few terms, though with increasing order its effect diminishes and may become negligible, whereas ρ^2 remains constant throughout. The field series converge more slowly than those for mutual inductance, because of the lower value of the exponent p, and for the same reason the series for filament systems converge more slowly than for solenoids, especially thick solenoids. But this advantage of thick solenoids is partially offset by a tendency to higher values of ρ. With massive windings it is hard to avoid sources near the origin in central

‖ ‖ The factor $1-(1/4m)$ of Eq. (13) has only a second-order effect.

TABLE VIII. Cotangents for fourth-order split solenoids.*

c_1	c_2	δ_m''	c_1	c_2	δ_m''	c_1	c_2	δ_m''	c_1	c_2	δ_m''	c_1	c_2	δ_m''
01	2.955 308		11	1.315 689	5 130	21	0.956 437	1 304	31	0.747 347	592	41	0.600 937	345
02	2.408 057		12	1.266 269	4 260	22	0.931 199	1 185	32	0.730 538	556	42	0.588 519	330
03	2.120 617	70 000	13	1.221 145	3 590	23	0.907 149	1 081	33	0.714 285	524	43	0.576 433	316
04	1.928 870	44 000	14	1.179 636	3 060	24	0.884 182	991	34	0.698 557	494	44	0.564 662	303
05	1.786 205	27 800	15	1.141 209	2 640	25	0.862 208	911	35	0.683 323	467	45	0.553 195	291
06	1.673 134	18 900	16	1.105 442	2 310	26	0.841 147	842	36	0.668 557	442	46	0.542 018	279
07	1.579 736	13 580	17	1.071 992	2 030	27	0.820 930	780	37	0.654 233	420	47	0.531 121	268
08	1.500 308	10 180	18	1.040 581	1 800	28	0.801 493	725	38	0.640 329	399	48	0.520 492	258
09	1.431 285	7 900	19	1.010 975	1 605	29	0.782 782	676	39	0.626 824	380	49	0.510 121	249
10	1.370 296	6 300	20	0.982 982	1 443	30	0.764 748	632	40	0.613 699	362	50	0.500 000	240

* The cotangents c_1 and c_2 are ratios of length to diameter for the central gap and for the winding. The column δ_m'' gives second central differences modified by the "Comrie throwback" to include the fourth order, i.e., $\delta_m'' = \delta'' - 0.184\delta^{IV}$. To interpolate f_θ between consecutive entries f_0 and f_1, where t is the fraction of the tabular interval 0.01 and $\phi = 1-\theta$, first compute an intermediate difference $\delta_\theta'' = \frac{1}{2}[(1+\phi)\delta_{m0}'' + (1+\theta)\delta_{m1}'']$. Then $f_\theta = \phi f_0 + \theta f_1 + \frac{1}{2}\phi\theta\delta_\theta''$. .

fields and undue projections upon search coils, in relation to the mean radius. This affects adversely both the size of the low error region and the space in which the field is easily calculated with precision. But if the range of central uniformity is appraised in relation to the inner radius, experience confirms the excellence of the thick coil when carefully designed. It confirms also the general advantage of solenoid forms, hitherto inadequately exploited.

In practice, thin-solenoid fields are conveniently calculated to one part in 10^3 and thick-solenoid fields to one part in 10^4, using five or at most six error terms, when ρ is as high as 0.7. Note that ρ is a ratio of polar radii, and it may sometimes be expedient to lower it by shifting the origin. With the same value 0.7, thin-solenoid mutual inductances can be found to better than one part in 10^6, using ten terms computed from Table III. The method is less arduous than those in common use for precise calculation of M. Fields may still be computed to one-tenth percent with ρ about 0.8. Higher values are increasingly awkward, and a zonal harmonic expansion seems incapable of yielding even an approximate result for self-inductance ($\rho = 1$).

SPECIAL CENTRAL-FIELD SYSTEMS

Here the methods and results of earlier sections are applied to certain special systems for illustration.

Filament Systems Approximating Harmonic Windings

Loop combinations that suppress several central error constants q_n immediately following q_1 or q_2 have been ably discussed by McKeehan.[16,22] Early in the present work, it was noticed that one of the four-loop systems[23] adds a fourth member to a series that began with Ampere's loop, the Helmholtz pair, and Maxwell's three-loop combination.[15] The series may be indefinitely extended; all of its members generate uniform H_z and may be designated as of harmonic type h_1 (Eq. (30)). To indicate this, only the subscript (1) will be retained, and the system will be classified as $(P_m')_1$, since the $(m-1)$th member consists of $m-1$ filaments placed on the surface of a sphere, at positions u_{jm} defined by the roots of $P_m' = 0$. The turns N_{jm} are proportional to a set of constants A_{jm}' from which further series $(P_m')_2$, $\cdots$, $(P_m')_n$ may be derived. $(P_m')_2$ generates a gradient (h_2) while $(P_m')_n$ generates a field of type h_n. The turn ratios of the general system conform to the equation,

$$N_{jmn} = A_{jm}' P_n'(u_j) \times \text{const}, \tag{59}$$

and occupy the same positions on the sphere for all values of n. Whether m is odd or even, the $(P_m')_1$ system has a lowest order error constant q_{2m-1}, hence its order of error is $2m-2$.

Of the gradient systems, only Maxwell's $(P_4' = 0)$ has been described previously. Systems of even m have errors of order $2m-4$, using only $m-2$ filaments, while odd-m systems are less economical in this respect. In even-m $(P_m')_1$ systems, beginning with Maxwell's, the equatorial coil corresponding to $P_m'(0) = 0$ has the largest turn count (maximum A_{jm}'). But in gradient systems, the equatorial coil is suppressed altogether since the factor P_n' of Eq. (59) becomes $P_2'(0)$, which vanishes.

For each series $(P_m')_n$ there is a corresponding series $(P_m)_n$, in which filaments of strength A_{jm} occupy positions at the m roots of $P_m = 0$. Only one member of any such series, namely $(P_4)_1$, has been previously noted.[24] They require one filament more than (P_n') systems, for the same order of error.¶¶

[22] L. W. McKeehan, Rev. Sci. Instr. 7, 178 (1936).

[23] Reference 16, Table III, Note (e). McKeehan seems not to have noticed that these loops coincide with the roots of $P_5' = 0$.

[24] Reference 16, Table III, Note (c).

¶¶ As stated above, Figs. 2 to 6 may represent electrostatic equipotentials, contours of constant axial electric field E_z, of constant gradient $E_z^{(1)}$, etc. The generating systems may be charged rings on a spherical surface, in the same positions occupied by the magnetic filaments, and may be classified into $(P_m)_0$, $(P_m)_1$, and $(P_m)_2$ systems, with corresponding P_m' forms. In the electrostatic case, P_n appears in place of the P_n' of Eqs. (43) and (59), while it is the (P_n) systems that require fewest rings for a given order of error.

All these systems evidently approximate the continuous spherical current sheets previously described. In fact, A_{jm} and A_{jm}' are just the weight coefficients that occur in the numerical integration formulas of Gauss and of Radau,[25] respectively. Each filament system stands in precisely the same relation to the true harmonic current sheet as does the weighted summation of ordinates to the integral for which it substitutes.*** Since the publication of 15-figure gaussian coefficients to $m=16$,[26] only common sense forbids the construction of a 16-filament system with 30th-order error!

The remote-field equivalents of Eqs. (43), (44) describe a set of constants p_n analogous to q_n. Those immediately following p_1 or p_2 determine the errors of search coil systems, for field or for gradient, and the equations show that whenever $q_n=0$ for a spherical filament system, p_n also vanishes. Thus every spherical system $(P_m')_n$ or $(P_m)_n$ may in theory serve as a search coil to measure the particular derivative $H_x^{(n)}$ that it generates as a current system. But the most practical search coils are thick solenoids. Excellent starting points for their design are the solenoid analogs of spherical filament systems. In these the filaments at the roots of $P_m'=0$ become end-turns of solenoids, whose turn-densities N' replace N in Eq. (59). If the filament system suppresses q_n, hence also p_n, its solenoid analog suppresses q_{n+1} (see Eqs. (43), (44)) and also p_{n-1}. The first member of the series derived from $(P_m')_2$ is Gray's solenoid, with ends in the Maxwell position; here q_3 and p_3 vanish. Further discussion of this series, including thick-solenoid systems derived from it, will be deferred to a later paper.

Fourth-Order Thin Solenoids

Equations (44) show that q_3 cannot vanish for a single solenoid unless $u=0=q_1$, or $\sin^2\alpha=0$; that is, the solenoid is either infinitely short or infinitely long. The second-order error of any finite solenoid may be corrected by omitting or short-circuiting some equatorial turns. The solenoid may then be regarded as an elongated Helmholtz pair whose cross section has an infinite aspect ratio. Such a system has the well-known precision of a cylindrical winding, with better field uniformity than a conventional Helmholtz pair of the same cylindrical radius a; also it is more easily cooled, and often it fits better into the available space.[27] It may be designed with a gap wide enough to admit a shaft or other device. Equation (44) shows that the positions of the ends are given by

$$u_1 \sin^4\alpha_1 = u_2 \sin^4\alpha_2. \tag{60}$$

In practice the cotangents x/a are more useful, and these have been computed for Table VIII at intervals of 0.01 over the complete range of $\cot\alpha_1$. Interpolation using the tabulated differences permits a precision in excess of any experimental requirement. Table IX lists the first two error coefficients, ϵ_4 and ϵ_6, for several of the solenoids of Table VIII. These correspond to a choice of $r_0=a$ in Eq. (31). For slight elongations of the ideal Helmholtz filaments ($c_1=c_2=0.5$) the improvement is of course not marked.

In the same way, and with similar advantages, Maxwell's gradient pair may be replaced by a split solenoid with opposed windings. The equation to be solved is

$$\sin^5\alpha_1(5u_1^2-1)=\sin^5\alpha_2(5u_2^2-1). \tag{61}$$

Limiting cases are a semi-infinite pair with near ends in the Helmholtz position, and of course the infinitely short pair or Maxwell filaments. Intermediate solutions are easily found.

Fourth-Order Thick Solenoids

These will be treated in a later paper, by expansion of filament and of thin solenoid forms, and by using the integrated coefficients of Eqs. (48)–(50). For fourth-order fields or gradients, the Helmholtz or Maxwell pair with general rectangular section meets all requirements, since these simplest systems afford only a single adjustable parameter, and one suffices to annul the second order. In more complex systems, additional adjustments are possible, which in a well-designed system ought to be used to suppress errors of higher order than the second. It matters little whether the axial spacing or one of the two section dimensions is chosen as the single disposable parameter. Two of these three may be set arbitrarily and when the third is adjusted to annul q_3 or q_4, the fourth-order error is nearly the same for any coil cross section. It will be shown that for square sections whose side A is no greater than four-fifths of the mean radius a_0, the proper Helmholtz spacing of section centers is, to a fraction of one percent,

$$2x=a_0[1+0.02083(A/a_0)^2+0.0703(A/a_0)^4] \tag{62}$$

and that the fourth-order error is nearly the same as for a filament pair of radius a_0 at its proper spacing a_0.

When the normal axial spacing of a thick Helmholtz pair has been slightly increased or decreased, fields

TABLE IX. Error coefficients ϵ_4 and ϵ_6 for split solenoids. (Referred to $r_0=a$.)

c_1	0.50	0.40	0.35	0.30	0.25	0.20	0.15	0.10	0.05
ϵ_4	−1.152	−1.098	−1.027	−0.925	−0.794	−0.640	−0.470	−0.297	−0.133
ϵ_6	+1.262	+1.191	+1.100	+0.972	+0.814	+0.638	+0.456	+0.282	+0.128

[25] R. Radau, J. Math. (3) 6, 307 (1880).

*** This elegant relation was pointed out to me by W. F. Brown (private communication) after I had spent much time computing A_{jm}' through $m=9$, and A_{jm} through $m=7$, by a semi-empirical method based on an elaboration of Eq. (64). (See M. W. Garrett, Phys. Rev. 74 (A), 1231 (1948).) The coefficients of Radau (to $m=10$) date back to 1880, and those of Gauss to 1814!

[26] Lowan, Davids, and Levenson, MT 18, Natl. Bur. Standards (1942).

[27] This convenient design seems to have escaped notice, except that A. and F. Sauter, Z. Physik 122, 120 (1944) solved three cases graphically to about five percent.

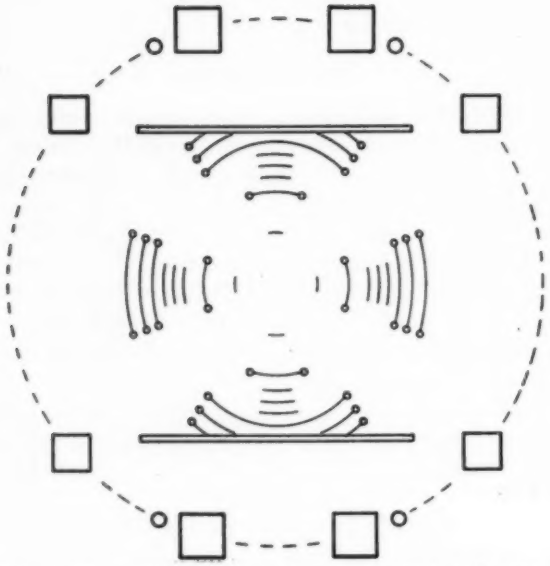


FIG. 7. Three eighth-order systems, with Helmholtz pair for comparison (small circles and innermost error limits). The long arcs with circle ends are 1 percent error limits, the short arcs are 0.1 percent limits.

identical with those of Fig. 5 or of Fig. 6 (respectively) result. The oblate field is useful for cloud-chamber coils, and examples of the arbitrary rectangular cross section will be calculated, including the errors of tenth order and lower. Here again, the section may be chosen freely, and the field regulated by varying the axial spacing alone.

Sixth-Order Solenoids

Many workers have employed solenoids with overwound ends but no one has made full use of the two available arbitrary constants. If a single-layer solenoid is double-wound at the outer ends, with a central gap, there exist unique values of both angles, i.e., that of the total length and that of the singly-wound gap, which annul both q_3 and q_5 of Eq. (51). The correct values are:

$$\cot\alpha_1=1.82234, \quad \cot\alpha_2=1.39327. \tag{63}$$

The first six error coefficients are, in parts per million choosing $r_0=a$ in Eq. (31), -5504 (6), -2634 (8), -193 (10), $+238$ (12), $+82$ (14), and -5 (16). This field will later be mapped in detail, since the improvement over other systems that can be precision-wound is drastic. H_z is uniform to one part in 10^3 out to $0.72a$ on the axis, and to $0.98a$ at the equator, while the axial 1 percent limit is $1.03a$.

Using this solenoid as a point of departure, it has been found possible to design sixth-order thick solenoids with nearly complete freedom in the choice of their dimensions, except that the innermost layer must have a length somewhat in excess of its diameter. The correction is effected by removal of turns in a single region of rectangular section. If considered in relation to the inner radius, thick solenoids are even better than thin. Extreme care will be required in winding if their excellent theoretical performance is to be realized.

Eighth-Order Systems

A single set of three systems will be discussed here (Fig. 7). The areas of the rectangular sections in the figure indicate the turn ratios correctly, but the calculations assume ideal filaments at the centers of the squares ($P_5'=0$) and a thin current sheet at the mean radius of the long coil of rectangular section ($P_4'=0$). The small circles mark the Helmholtz positions, for reference only. The three eighth-order fields are produced by combining the two (square) loop-pairs, or by the use of either pair with the solenoid. The four-loop system is McKeehan's.[23] It was later rediscovered by Sauter and Sauter.[27] The following derivation is both simpler and more general than those previously given. From Eq. (11), if

$$(P_n'=0), \text{ then, } P_{n+1}'/P_{n-1}'=-(n+1)/n$$

$$P_{n+2}'/P_{n-2}'=-(2n+3)/(2n-1). \tag{64}$$

It follows from this equation and (43) that, for either of the pairs of filaments at the roots of $P_5'=0$,

$$P_6'/P_4'=-6/5, \quad q_7/q_3=-13/9. \tag{65}$$

It is easily verified from Eqs. (43) and (1) that when both filament pairs are used, with a turn ratio 1.4660372 (as shown in Fig. 7) q_3 for the combination vanishes. But (65) shows that the same ratio of turns that annuls q_3 must also annul q_7, and since q_5 vanishes for each loop pair separately, an eighth-order field results. When the remote field is discussed, it will appear that the left-hand equation of (65) leads in just the same way to a sixth-order search-coil system composed of two solenoids. The same idea, when applied to the roots of $P_4=0$, using Eq. (9) in place of (11), leads to McKeehan's $(P_4)_1$ combination.[24]

Either pair of P_5' loops may be combined with Gray's solenoid ($P_4'=0$) to yield an eighth-order field. It appears from Eq. (44) that for this third element also $q_5=0$, while the solenoid and the loop pair may be made to annul q_3 and q_7 simultaneously since there remain two adjustable ratios; that of the total solenoid turns to total loop turns (N_s/N_f) and that of the polar radius of the solenoid to that of the loop pair (r_s/r_f). Set $q_n(f)+q_n(s)=0$ $(n=3, 7)$ substitute values of q_n from (43) and (44), and divide the equation in q_7 by that in q_3. The result is:

$$(r_s/r_f)^4=-\frac{1}{3}\frac{P_3'(u_f)P_6'(u_s)}{P_7'(u_f)P_2'(u_s)}=\frac{1}{3}\times\frac{-9}{13}\times\frac{-11}{7}=\frac{33}{91} \tag{66}$$

where both P' ratios have been evaluated from (64). The same ratio r_s/r_f holds for either filament pair. Since the solution is now easily completed, further algebraic details are omitted.

The signs both of q_3 and of q_7 are opposite for the paraxial and equatorial filament pairs, otherwise one current would have to be reversed in the original $(P_5')_1$ combination. It is found that the solenoid errors are of the right sign to annul those of the paraxial pair, and the field of this three-element combination is more uniform than that of any other known system of comparable simplicity. When the equatorial pair is used with Gray's solenoid, one of the currents must be reversed. Such a combination, besides having lower current efficiency, yields a less uniform field, but it might prove useful in some application that demands a highly oblate over-all geometry, or an extremely wide angle of access to the central region.

In the figure, arcs ending in small circles indicate 1 percent error limits of various systems, while the short plain arcs are 0.1 percent limits. The innermost limits, given for comparison, belong to the Helmholtz pair. The wider limits occur in sets of three. Of these the outermost correspond to the optimum combination of Gray's solenoid with the paraxial P_5' pair. These enclose the limits of the $(P_5')_1$ loop system, which in turn enclose those of the third system using opposed currents. The optimum system might be mounted quite simply on a single cylinder. Taking into account the difficulty of extending the inherently sharp boundary of the eighth-order field, it may be seen that the expansion shown in Fig. 7 represents a significant advance in design and corresponds to a very low coefficient of error. Relative numerical data are assembled in Table X. The total turns N_s, polar radius r_s, and central field H_z of the Gray's solenoid are arbitrarily taken as unity. Errors are those of ϵ_8 alone, which is here referred to $r_0=r_s$.

TABLE X. Eighth-order field systems.

All dimensions are referred to r_s					
Element	N (total)	r	u	a	H_z
Solenoid	1.000	1.000	0.6546537	0.755929	1.000
Paraxial pair	1.530572	1.288641	0.7650553	0.829839	0.4925
Equatorial pair	−2.243876	1.288641	0.2852315	1.235109	−1.600

		Axial limits		Equatorial limits	
Combination	ϵ_8	0.1 percent	1 percent	0.1 percent	1 percent
Solenoid-paraxial	0.141	0.54	0.72	0.63	Wall
Four loops	0.297	0.49	0.65	0.58	0.77
Solenoid-equatorial	−0.683	0.44	0.59	0.52	0.69

It is a pleasure to acknowledge the assistance of Alfred Curtis Hunting and of Laura C. Maurer, who carried out many of the calculations for Figs. 5 and 6, Table VIII, and the coefficients A_{jm}' and A_{jm}, which were later found to check accurately with those of Radau and of Gauss.

Information Theory and Most Efficient Codings for Communication or Memory Devices

L. BRILLOUIN
International Business Machines Corporation, New York, New York
(Received February 12, 1951)

Shannon's theorem about the capacity of a channel is discussed, and it is shown that the most efficient coding is the one yielding the most probable distribution of the code symbols. A specific rule is obtained for this most probable distribution. Most efficient coding is essential for communication channels or memory devices in large scale computers.

I. INTRODUCTION

IN a recent paper,[1] the author discussed the connection between entropy and information, and used a statistical method to justify Shannon's fundamental formula about information.[2] The aim of the present paper is to apply similar methods to the definition of the "capacity of a channel" according to Shannon, and to discuss Shannon's great theorem about codings, enabling one to use a channel to its full capacity. It will be shown that the best codings are the ones yielding the most probable distribution of the symbols, and specific rules will be given for finding these optimum coding systems.

Let us first rewrite here some of the results of the preceding paper.[1] We assumed that we were using n different symbols in our communication system, and we computed the total number of different combinations of such symbols (let us call them "telegrams") when the numbers $N_1N_2\cdots N_j\cdots N_n$ of the different symbols used in the telegram were given in advance. Let

$$G_0=N_1+N_2\cdots+N_n=\sum_j N_j \tag{1}$$

be the total number of symbols in the telegram and let

$$p_j=N_j/G_0, \quad \sum_{j=1}^{n} p_j=1 \tag{2}$$

represent the relative density of the jth symbols. The total number of different telegrams of this type is given by (reference 1, Eq. (17))

$$P=G_0!/\Pi_j N_j! \tag{3}$$

or approximately, using Stirling's formula,

$$\ln P=-G_0\sum_j p_j \ln p_j \tag{4}$$

(reference 1, Eq. (43) with all the P_j's equal to unity). This formula was used to define the amount of information obtained when one special telegram is selected out of the P possibilities:[3]

$$I=-kG_0\sum_j p_j \ln p_j, \tag{5}$$

where k represents Boltzmann's constant.

[1] L. Brillouin, J. Appl. Phys. **22**, 338 (1951).
[2] C. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, Illinois, 1949).
[3] Reference 1, Eq. (47).

II. WHAT DO WE CALL A "CHANNEL"?

If we want to transmit information or to store it in a memory device, we first select a *physical medium:* a cable, a radio link, a mercury line, a magnetic tape, etc., and we have to study the limitations occurring in this physical problem. For instance, if we use a cable with a band width $\Delta\nu$, the shortest signals to be transmitted will have a duration Δt:

$$\Delta t\cdot\Delta\nu=1. \tag{6}$$

Next comes a technical problem of devising a set of symbols, adapted to the physical medium, and enabling us to use it for communication; dots, dashes and spaces, signals of different intensities or polarity, etc. Once this choice is made, we have defined a *communication channel.* For our statistical discussion, the main point is the set of n discrete symbols, $S_1\cdots S_j\cdots S_n$ to be used, and the length (or duration) $t_1\cdots t_j\cdots t_n$ of these symbols.

Shannon[4] defines the *capacity of a discrete channel* in the following way: He considers a long interval of time T and computes the total number $N(T)$ of different telegrams of total duration T that can be obtained by using all the available symbols in all possible ways (instead of "telegram," Shannon uses the word "signal" which may lead to misunderstandings). The capacity of the channel is defined as

$$C=(k/T)\operatorname*{Lim}_{T\to\infty}\ln N(T). \tag{7}$$

If logarithms of base 2 are used instead of napierian logarithms, the constant k should be replaced by

$$k'=k\ln 2, \tag{8}$$

and the capacity will be given in *bits per second.* The constant k is needed if we want to measure information in usual entropy units.

Let us first assume that no restrictions or constraints are applied on the possible sequences of symbols. Let $N(t)$ represent the number of sequences of total duration t, we have a relation,[5]

$$N(t)=N(t-t_1)+N(t-t_2)\cdots+N(t-t_n). \tag{9}$$

It simply states that the total number $N(t)$ is equal to the sum of sequences ending in symbols, S_1, S_2, or S_n and these are $N(t-t_1)\cdots N(t-t_n)$, respectively.

[4] Reference 2, p. 7.
[5] Reference 2, p. 5.

It is a well-known result of the theory of finite differences that for very large t the function $N(t)$ will be asymptotic to AX^t and increase exponentially. Substituting this asymptotic expression in Eq. (9), we have

$$X^t = X^{t-t_1} + X^{t-t_2} \cdots X^{t-t_n},$$

or

$$X^{-t_1} + X^{-t_2} \cdots + X^{-t_n} = 1. \tag{10}$$

We must choose the largest real solution X_0 of this characteristic equation (10) and we obtain $N(t) \rightarrow AX_0^t$; hence,

$$C = k \ln X_0. \tag{11}$$

Shannon also discusses problems with restrictions imposed on the possible sequences of symbols S_j, but we shall restrict our present discussion to the preceding problem with no restrictions or constraints.

After defining the capacity of a channel, Shannon introduces his definition of the entropy of information, which we discussed in a previous paper. We called i the density of information per cell,[6] which is thus defined in bits per symbol.

Shannon proves a fundamental theorem[7] for a *noiseless channel:*

If the channel has a capacity C (bits per second) and accepts telegrams from a source of information i (bits per symbol), then the best possible coding system will enable us to use the channel at the rate of $R = C/i$ (symbols per second).

We intend to discuss more precisely the physical meaning of this very important theorem, and to specify the properties of the coding systems enabling one to use the channel to full capacity.

III. A FIRST EXAMPLE: SYMBOLS OF EQUAL LENGTH

Let us consider first a simple example, assuming all symbols to have an equal length t_1. A long time interval T contains

$$G = T/t_1 \tag{12}$$

positions (or cells) for different signals. For each cell we can use n different symbols, hence

$$N(T) = n^G = n^{T/t_1} \tag{13}$$

a result which checks with previous computations.[8] Our definition (7) now yields for the channel capacity,

$$C = (k/t_1) \ln n. \tag{14}$$

Shannon considered a special example of this type, when he computed the channel capacity for teletype symbols. In the problem of channel capacity, no assumption is made about the *a priori* density of the n different symbols. When we defined "information," according to Shannon, we assumed that we knew the *a priori* densities of the different symbols: p_1, $p_2 \cdots p_j \cdots p_n$ and we obtained for the information density (per symbol):

$$i = -k \sum_{j=1}^{j=n} p_j \ln p_j \tag{15}$$

as stated in Eq. (5). The information density i is maximum when all p_j's are equal

$$p_1 = \cdots p_j = \cdots p_n = 1/n;$$

hence

$$i = -k \sum_{i=1}^{n} (1/n) \ln(1/n) = k \ln n, \tag{16}$$

since we have n equal terms in the sum $\sum$. This distribution is the most *probable distribution*, and yields the largest possible number of different symbol distributions (or telegrams) when the total number G of symbols is given.

The similarity between Eqs. (14) and (16) is striking. In the computation of channel capacity we used all possible distributions of symbols. In the case of information, we assumed an arbitrary *a priori* set of densities, which we adjusted later on (taking $p_j = 1/n$) in order to maximize the probability, and we end up with identical results.

When averaging over a very long period of time, the average distribution becomes identical with the most probable distribution. In other words, making no assumption about *a priori* densities is equivalent to assuming the most probable density choice.

With these general remarks we may now discuss Shannon's great theorem and specify that the *most efficient coding is the one yielding for the different symbols the most probable a priori density distribution.*

This is almost obvious now, since Shannon's relation yields a rate of transmission,

$$R = C/i = 1/t_1 \tag{17}$$

from Eqs. (14) and (16) and we certainly cannot transmit more than $1/t_1$ symbols of length t_1 per second. Hence, in our simplified problem, the best coding is the one using all symbols in equal proportion.

IV. GENERAL PROBLEM: SYMBOLS WITH DIFFERENT LENGTHS

We now want to discuss the general problem of Sec. II, with symbols of different lengths, and we shall prove that our preceding result still applies:

"The most efficient coding is the one yielding for the different symbols the most probable *a priori* density distribution." Furthermore, we are going to define this most probable distribution and to explain the meaning of our previous Eq. (10).

The discussion is slightly more involved for symbols of different lengths, and can be made in the following way. We start with the assumption of a telegram of total duration T, and we first specify that we are using $N_1, N_2, \cdots N_j \cdots N_n$ symbols of types $1, 2 \cdots j \cdots n$. The total

[6] Reference 1, Eqs. (36) and (47).
[7] Reference 2, Sec. 9, p. 28.
[8] Reference 1, Eq. (24).

duration of the telegram is

$$T=\sum_{j=1}^{n} N_j t_j. \tag{18}$$

We introduce the total number of symbols G_0 and their relative proportions p_j

$$G_0=\sum_{j=1}^{n} N_j, \quad p_j=N_j/G_0, \tag{19}$$

hence,

$$\sum_{j=1}^{n} p_j=1 \tag{20}$$

and

$$T=G_0\sum_{j=1} p_j t_j. \tag{21}$$

The total number P of different telegrams that we can build by interchanging the positions of the different symbols is still given by Eqs. (3) and (4),

$$\ln P=-G_0\sum_{j=1}^{n} p_j \ln p_j, \tag{22}$$

and all these telegrams have the same duration T, since they use the same numbers $N_1\cdots N_n$ of the different symbols.

Now we want to obtain the *most probable distribution*. This means maximizing $\ln P$, while taking into account the two conditions (20) and (21). This is done with the method of Lagrange's multipliers. We simply write the condition,

$$d(\ln P)-\alpha d(\sum p_j)-\beta dT=0, \tag{23}$$

with two arbitrary multipliers α and β to be determined later. The variables are G_0, $p_1\cdots p_j\cdots p_n$. Differentiating (22), we obtain

$$d(\ln P)=-dG_0\sum p_j \ln p_j-G_0\sum dp_j(1+\ln p_j).$$

We use this expression in condition (23) and obtain

$$-dG_0[\sum p_j \ln p_j+\beta\sum p_j t_j]$$
$$+\sum dp_j[-\alpha-\beta G_0 t_j-G_0(1+\ln p_j)]=0. \tag{24}$$

We must now treat the dG_0 and dp_j's as independent quantities. The total differential can be zero only when all brackets are zero:

$$\sum p_j(\ln p_j+\beta t_j)=0, \tag{25}$$

$$\ln p_j=-1-\beta t_j-\alpha/G_0, \quad j=1, 2, \cdots n. \tag{26}$$

Equation (26) can be used in (25) and gives

$$\sum p_j(-1-\alpha/G_0)=0, \tag{27}$$

but we also have condition (20); hence, we must choose

$$1+\alpha/G_0=0, \quad \alpha=-G_0 \tag{28}$$

and (26) reduces to

$$\ln p_j=-\beta t_j, \quad p_j=e^{-\beta t_j}. \tag{29}$$

The arbitrary constant β is determined by condition (20):

$$\sum_{j=1}^{n} p_j=\sum_{j=1}^{n} e^{-\beta t_j}=1; \tag{30}$$

but if we write

$$X=e^{\beta}, \tag{31}$$

we note that this equation is identical with Shannon's Eq. (10) of Sec. II. We have again the identity between the average distribution (Shannon) and the most probable one (that we have just computed) when we consider infinitely long telegrams.

Let us now complete the discussion. Our most probable distribution yields a number P of different telegrams (Eq. 22):

$$\ln P=-G_0\sum_j p_j \ln p_j=+G_0\sum_j \beta t_j e^{-\beta t_j}=G_0\beta\sum t_j p_j$$

or

$$\ln P=\beta T, \quad P=e^{\beta T}=X^T, \tag{32}$$

according to condition (21). Our P (most probable distribution) is again identical with Shannon's $N(T)$ corresponding to all possible distributions, and the channel capacity, according to formula (11) is

$$C=k\ln X=k\beta. \tag{33}$$

For long telegrams, the most probable distribution is so much more probable than any other one that all other possible distributions can be simply omitted and ignored. The most efficient coding is the one that obtains this most probable distribution of symbols, namely,

$$(\ln p_j)/t_j=-\beta \text{ a constant} \tag{34}$$

according to (29). The largest real root β of Eq. (30) corresponds to the largest X solution of Shannon's Eq. (10) and yields, according to (32), the utmost probable distribution, with the largest possible P value.

If we use the most efficient code, and a distribution of symbols corresponding to the most probable one, we have a density of information per symbol

$$i=-k\sum_j p_j \ln p_j=+k\beta\sum_i t_j p_j=+k\beta\bar{t} \tag{35}$$

calling $\bar{t}$ the average length of a symbol, when the probability of the different symbols is (29) or (34) and the rate of transmission in (Eqs. (33), (35)), is

$$R=C/i=1/\bar{t}, \tag{36}$$

a formula very similar to (17).

The smaller the probability p_j of a certain symbol S_j, the larger $(-\ln p_j)$. Hence, the longer the symbol duration t_j according to Eq. (34). This explains the exact meaning of this relation.

V. CONCLUSION

The condition (29) or (34) corresponds to the coding system that matches the channel (or the memory system) most efficiently. It is just as important in coding problems, as the condition of matching characteristic

impedances, in circuit problems. Shannon uses, in a special problem, a procedure that leads to a very similar rule.[9] We proved here that the result is absolutely general and applies to all coding problems.

Let us note, in concluding, that the coincidence of "average" and "most probable" distribution is a well-known result for extremely large numbers of events. This rule plays a most important role in statistical thermodynamics, where it explains the coincidence between different definitions given by J. W. Gibbs, Maxwell, Boltzmann, or Planck. This point was very clearly discussed by H. A. Lorentz.[10]

[9] Reference 2, pp. 29–30.

Note added in proof.—The problem of Maxwell's demon was very carefully discussed by P. Demers in two interesting papers. [Can. J. Research **22**, 27 (1944) and **23**, 47 (1945)]. The author analyzes previous discussions presented by Lewis, Slater, and Smoluchowski and comes to final conclusions that are very similar to those of the present author. [J. Appl. Phys. **22**, 334 (1951)].

[10] H. A. Lorentz, *Leçons au College de France* (*Les theories statistiques en Thermodynamique* (B. G. Teubner, Leipzig, 1916)).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 9 SEPTEMBER, 1951

Theory of V-Antennas

RONOLD KING
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received March 14, 1951)

An integral equation for the current in an apex-driven symmetrical V-antenna is derived and solved by successive approximations. General formulas for the distribution of current and the impedance are obtained. The zeroth-order impedance is evaluated for a leg-length $h=\lambda_0/4$ as a function of the enclosed angle and used to estimate experimental results.

1. INTEGRAL EQUATION FOR SYMMETRICAL V-ANTENNA

THE symmetrical, center-driven V-antenna shown in Fig. 1 is a general form of the symmetrical, center-driven antenna. It differs from this only in having the identical halves of the antenna inclined at an arbitrary angle Δ with respect to each other instead of being collinear with $\Delta=\pi$. Let half No. 1 of the antenna lie along the z-axis; the other half, No. 2, along the s-axis in the xz-plane with positive direction defined by the unit vector

$$\boldsymbol{s}=\boldsymbol{z}\cos\Delta+\boldsymbol{x}\sin\Delta. \tag{1.1}$$

A scalar potential difference,

$$V_\delta=\phi_1(z=\delta)-\phi_2(s=\delta)=2\phi_1(\delta), \tag{1.2}$$

is maintained across the base of the V-structure by a balanced transmission line so that the currents and charges in the halves satisfy the conditions,

$$I_{2s}(s')=-I_{1z}(z'); \quad q_2(s')=-q_1(z'), \tag{1.3}$$

when s' equals z'. Terminal-zone effects and antenna-to-line coupling effects are ignored in the analysis. Account may be taken of them by as uitably designed lumped-constant network.

The derivation of the integral equation for the current in half No. 1 of the V-antenna parallels the derivation for the antenna with collinear halves[1] to which the V-antenna reduces when $\Delta=\pi$. For simplicity, it is assumed that the two conductors have the same radius a and are perfect, so that the tangential components of the electric field vanish on their surfaces. If required, terms to take account of ohmic resistance may be added. The boundary condition at a point P on conductor No. 1 is

$$\left.\begin{matrix} r_1=a \\ \delta\leqq z\leqq h \end{matrix}\right\}: E_{1z}=-(\partial\phi/\partial z)_1-j\omega A_{1z}$$
$$=-j(\omega/\beta_0^2)((\partial/\partial z)\,\mathrm{div}\boldsymbol{A}_1+\beta^2{}_0A_{1z})\doteq 0. \tag{1.4}$$

Since with $A_y=0$,

$$\mathrm{div}\boldsymbol{A}_1=(\partial A_{1z}/\partial z)+\partial A_{1x}/\partial x, \tag{1.5}$$

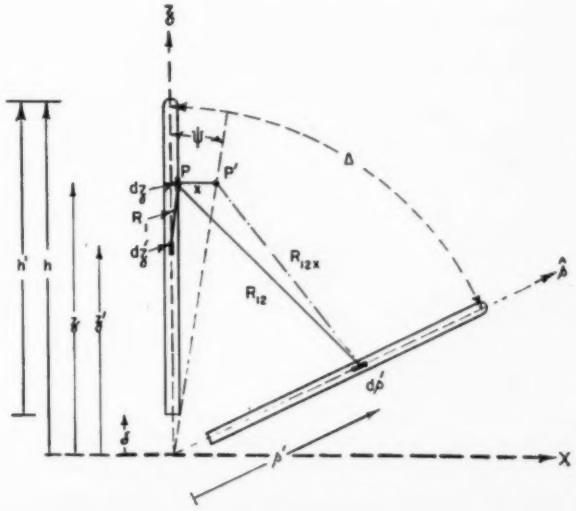


FIG. 1. V-antenna.

[1] R. King and D. Middleton, Quart. Appl. Math. **3**, 302 (January, 1946).

Eq. (1.4) yields the following differential equation valid on the surface of conductor No. 1:

$$(\partial^2 A_{1z}/\partial z^2)+\beta_0^2 A_{1z}=-\partial^2 A_{1x}/\partial z\partial x;\ \left\{\begin{matrix} r_1=a \\ \delta\leqq z\leqq h\end{matrix}\right\}. \quad (1.6)$$

Expressions for the components of the vector potential in (1.6) are:

$$A_{1z}=\frac{1}{4\pi\nu_0}\int_\delta^h\left\{I_{1z}(z')\frac{e^{-j\beta_0 R_1}}{R_1}dz' + \cos\Delta I_{2z}(s')\frac{e^{-j\beta_0 R_{12}}}{R_{12}}ds'\right\}, \quad (1.7a)$$

$$A_{1z}=\frac{\sin\Delta}{4\pi\nu_0}\int_\delta^h I_{2z}(s')\frac{e^{-j\beta_0 R_{12}}}{R_{12}}ds', \quad (1.7b)$$

where $\nu_0=1/\mu_0=10^7/4\pi$ meters/henry and

$R_1\equiv((z-z')^2+a^2)^{\frac{1}{2}}$;

$$R_{12}\equiv(z^2-2zs'\cos\Delta+s'^2+a^2)^{\frac{1}{2}}. \quad (1.8)$$

The distance R_{12} defined in Eq. (1.8) is measured from ds' on the axis of conductor 2 to dz at P on the surface of conductor 1. It is assumed that P has the coordinates, $(x=0, y=a, z)$. In order to obtain $\partial A_{1z}/\partial x$ at $x=0$, it is necessary to evaluate $\partial A_{1z}/\partial x$ at some point P' at a distance x from P before x is set equal to zero. That is, it is necessary to use (1.7b) with

$$R_{12x}=(z^2+x^2-2(z^2+x^2)^{\frac{1}{2}}s'\cos(\Delta-\psi)+s'^2+a^2)^{\frac{1}{2}} \equiv(\xi^2+a^2)^{\frac{1}{2}}, \quad (1.9)$$

instead of R_{12} as given in (1.8). Note that

$$\cos(\Delta-\psi)=[z\cos\Delta/(z^2+x^2)^{\frac{1}{2}}]+x\sin\Delta/(z^2+x^2)^{\frac{1}{2}}, \quad (1.10)$$

so that

$$\xi^2=z^2+x^2-2zs'\cos\Delta-2xs'\sin\Delta+s'^2, \quad (1.11)$$

and

$$(d\xi/dx)_{x=0}=-s'\sin\Delta/(z^2-2zs'\cos\Delta+s'^2)^{\frac{1}{2}}. \quad (1.12)$$

With (1.9) to (1.11) the derivative of (1.7b) becomes

$$\left(\frac{\partial A_{1z}}{\partial x}\right)_{x=0}=\frac{\sin\Delta}{4\pi\nu_0}\int_\delta^h I_{2z}(s')\times\left[\frac{\partial}{\partial\xi}\left(\frac{e^{-j\beta_0 R_{12x}}}{R_{12x}}\right)\frac{d\xi}{dx}\right]_{x=0}ds'. \quad (1.13)$$

With (1.12) this becomes

$$\left(\frac{\partial A_{1z}}{\partial x}\right)_{x=0}=-\frac{\sin^2\Delta}{4\pi\nu_0}\int_\delta^h I_{2z}(s')\times\left[\frac{1}{\xi}\frac{\partial}{\partial\xi}\frac{\exp[-j\beta_0(\xi^2+a^2)^{\frac{1}{2}}]}{(\xi^2+a^2)^{\frac{1}{2}}}\right]_{x=0}s'ds'. \quad (1.14)$$

Let

$$\bar{K}_{12}(z,s')\equiv\left[\frac{1}{\beta_0\xi}\frac{\partial}{\partial\xi}\frac{\exp[-j\beta_0(\xi^2+a^2)^{\frac{1}{2}}]}{\beta_0^2(\xi^2+a^2)^{\frac{1}{2}}}\right]_{x=0} =-\frac{e^{-j\beta_0 R_{12}}}{\beta_0^3 R_{12}^3}(1+j\beta_0 R_{12}). \quad (1.15)$$

Also, let the symbol used for the variable of integration in (1.14) be changed from s' to z' after use has been made of (1.3a). This gives

$$\phi_{1x}(z)\equiv\frac{j\omega}{\beta_0^2}\left(\frac{\partial A_{1x}}{\partial x}\right)_{x=0}=\frac{j\zeta_0}{4\pi}\sin^2\Delta p(z), \quad (1.16a)$$

where

$$p(z)\equiv\beta_0^2\int_\delta^h I_{1z}(z')\bar{K}_{12}(z,z')z'dz', \quad (1.16b)$$

where $\bar{K}_{12}(z,z')$ is given by (1.15) with

$$R_{12}=(z^2-2zz'\cos\Delta+z'^2+a^2)^{\frac{1}{2}}. \quad (1.17)$$

With (1.16) the differential Eq. (1.6) becomes

$$(\partial^2 A_{1z}/\partial z^2)+\beta_0^2 A_{1z} =(j\beta_0^2/\omega)(\partial\phi_{1z}/\partial z);\ \left\{\begin{matrix} r_1=a \\ \delta\leqq z\leqq h\end{matrix}\right\}. \quad (1.18)$$

By differentiating (1.18) with respect to z and using the Lorentz condition,

$$(\partial A_{1z}/\partial x)+(\partial A_{1z}/\partial z)+j\beta_0^2/\omega\phi_1=0, \quad (1.19)$$

the following equation may be obtained for the scalar potential:

$$(\partial^2\phi_1/\partial z^2)+\beta_0^2\phi_1=\beta_0^2\phi_{1z}\left\{\begin{matrix} r_1=a \\ \delta\leqq z\leqq h\end{matrix}\right\}. \quad (1.20)$$

The solution of the nonhomogeneous differential Eq. (1.18) consists of the sum of a particular integral and a complementary function as follows:

$$A_{1z}=\frac{-j}{v_0}[C_1\cos\beta_0 z +C_2\sin\beta_0 z-\vartheta(z)];\ \left(\begin{matrix} r_1=a \\ \delta\leqq z\leqq h\end{matrix}\right). \quad (1.21)$$

This expression differs from the corresponding one for collinear halves only in the added particular integral

$$\vartheta(z)\equiv\int_\delta^z\frac{\partial\phi_{1z}(w)}{\partial w}\sin\beta_0(z-w)dw \quad (1.22)$$

and in the absence of the small ohmic terms. When $\Delta=\pi$, it follows from (1.16) that $\phi_{1z}(w)=0$. The factor $-j/v_0$ is introduced in (1.21) to make the constants and $\vartheta(z)$ dimensionally voltages; $v_0=3\times10^8$ m/sec.

The scalar potential on conductor 1 is obtained by

differentiating (1.21) using,

$$\phi_1=(j\omega/\beta_0^2)[(\partial A_z/\partial z)+\partial A_z/\partial x] =(j\omega/\beta_0^2)[\partial A_{1z}/\partial z]+\phi_{1x}. \quad (1.23)$$

Differentiation of (1.21) gives

$$\phi_1(z)=-C_1\sin\beta_0 z+C_2\cos\beta_0 z-\vartheta_1(z)+\phi_{1x}(z), \quad (1.24)$$

where

$$\vartheta_1(z)=\int_\delta^z(\partial\phi_{1x}(w)/\partial w)\cos\beta_0(z-w)dw. \quad (1.25)$$

The driving voltage V_δ is maintained across the gap from $z=\delta$ to $s=\delta$ by the transmission line. Using (1.2), it follows that

$$V_\delta=2\phi_1(\delta)=2[-C_1\sin\beta_0\delta+C_2\cos\beta_0\delta+\phi_{1x}(\delta)] \quad (1.26)$$

so that

$$C_2=[\tfrac{1}{2}V_\delta+C_1\sin\beta_0\delta-\phi_{1x}(\delta)]/\cos\beta_0\delta. \quad (1.27)$$

For convenience let

$$\tfrac{1}{2}\bar{V}_\delta\equiv\tfrac{1}{2}V_\delta-\phi_{1x}(\delta), \quad (1.28)$$

so that

$$C_2=[\tfrac{1}{2}\bar{K}_\delta+C_1\sin\beta_0\delta]/\cos\beta_0\delta. \quad (1.29)$$

This is like the corresponding formula for the straight cylindrical antenna but with $\bar{V}_\delta$ replacing V_δ.

The integral equation for the current $I_{1z}(z)$ is obtained by equating (1.21) with (1.29) to (1.7a). It is convenient to apply (1.3) and to change the symbol of the variable of integration from s' to z' in the second integral. The resulting integral equation is

$$\int_\delta^h I_{1z}(z')K_v(z,z')dz'=\frac{-j4\pi}{\zeta_0\cos\beta_0\delta}[C_1\cos\beta_0(z-\delta) +\tfrac{1}{2}\bar{V}_\delta\sin\beta_0 z]+\frac{j4\pi}{\zeta_0}\vartheta(z) \quad (r_1=a;\ \delta\leqq z\leqq h). \quad (1.30)$$

This may be arranged as follows:

$$\int_\delta^h I_{1z}(z')K_v(z,z')dz'=\frac{-j4\pi}{\zeta_0\cos\beta_0\delta}\{C_1\cos\beta_0(z-\delta) +\tfrac{1}{2}\bar{V}_\delta\sin\beta_0 z\}+J_v(z) \quad (r_1=a;\ \delta\leqq z\leqq h), \quad (1.31)$$

where

$$J_v(z)\equiv\frac{j4\pi}{\zeta_0}\left\{\int_\delta^z\frac{\partial\phi_{1x}(w)}{\partial w}\sin\beta_0(z-w)dw +\phi_{1x}(\delta)\frac{\sin\beta_0 z}{\cos\beta_0\delta}\right\}. \quad (1.32)$$

Integration by parts in (1.32) and rearrangement of terms using (1.16b) yields

$$J_v(z)=-\sin^2\Delta\left\{p(\delta)\tan\beta_0\delta\cos\beta_0(z-\delta) +\beta_0\int_\delta^z p(w)\cos\beta_0(z-w)dw\right\}. \quad (1.33)$$

The kernel in (1.31) is defined as follows:

$$K_v(z,z')\equiv(e^{-j\beta_0R_1}/R_1)-\cos\Delta e^{-j\beta_0R_{12}}/R_{12}. \quad (1.34)$$

Note that when $\Delta=\pi$, (1.31) reduces to the integral equation for the perfectly conducting cylindrical antenna. When $\Delta\neq\pi$, (1.31) is formally like the equation for the cylindrical antenna but with a different kernel and with the added term $J_v(z)$.

2. SOLUTION OF INTEGRAL EQUATION FOR V-ANTENNA

The integral equation (1.31) may be solved by the same method of iteration employed for the straight antenna ($\Delta=\pi$). Let an expansion parameter Ψ_v be defined as follows:

$$\Psi_v\equiv|\Psi_v(z_r)|\ ;\ \Psi_v(z)=\int^h g(z,z')K_v(z,z')dz', \quad (2.1)$$

where $g(z,z')$ is a distribution function that is chosen to make $I_{1z}(z)\Psi_v(z)$ approximate closely the integral on the left in (1.31). For this purpose the zeroth-order distribution is satisfactory, *viz.*,

$$g(z,z')=I(z')/I(z)=\sin\beta_0(h-z')/\sin\beta_0(h-z). \quad (2.2)$$

The resulting expression for the current is like that for the cylindrical antenna with appropriate changes in the functions as indicated by the addition of subscripts v.

$$I_{1z}(z)=\frac{j2\pi V_\delta}{\zeta_0\Psi_v}\times\left\{\frac{\sin\beta_0(h-z)+[M_{v1}{}^\delta(z)/\Psi_v]+[M_{v2}{}^\delta(z)/\Psi_v{}^2]+\cdots}{\cos\beta_0(h-\delta)+(A_{v1}{}^\delta/\Psi_v)+(A_{v2}{}^\delta/\Psi_v{}^2)+\cdots}\right\}. \quad (2.3)$$

In the following, only the first-order solution is considered. Higher order terms may be evaluated if required.

The first-order functions $M_{v1}{}^\delta(z)$ and $A_{v1}{}^\delta$ in (2.3) are defined as follows:

$$M_{v1}{}^\delta(z)=F_{v1}{}^\delta(z)\sin\beta_0 h-G_{v1}{}^\delta(z)\cos\beta_0 h +G_{v1}{}^\delta(h)\cos\beta_0 z-F_{v1}{}^\delta(h)\sin\beta_0 z, \quad (2.4a)$$

$$A_{v1}{}^\delta=F_{v1}{}^\delta(h)\cos\beta_0\delta+G_{v1}{}^\delta(h)\sin\beta_0\delta, \quad (2.4b)$$

where with $F_{0z}\equiv(\cos\beta_0 z-\cos\beta_0 h)$, $G_{0z}\equiv(\sin\beta_0 z-\sin\beta_0 h)$,

$$F_{v1}(z)=F_{0z}\Psi_v-\int_\delta^h F_{0z'}K_v(z,z')dz'+J_v(F_{0z},z), \quad (2.5a)$$

$$F_{v1}(h)=-\int_\delta^h F_{0z'}K_v(h,z')dz'+J_v(F_{0z},h), \quad (2.5b)$$

$$G_{v1}(z)=G_{0z}\Psi_v-\int_\delta^h G_{0z'}K_v(z,z')dz'+J_v(G_{0z},z), \quad (2.5c)$$

$$G_{v1}(h)=-\int_\delta^h G_{0z'}K_v(h,z')dz'+J_v(G_{0z},h), \quad (2.5d)$$

where $J_v(F_{0z}, z)$ is the same as $J_v(z)$ in (1.33) but with $F_{0z'}$ substituted for $I_{12}(z')$ in p. Specifically,

$$p(F_{0z}, z)=\beta_0^2\int_\delta^h F_{0z'}\bar{K}_{12}(z, z')z'dz'. \tag{2.6}$$

A corresponding function is defined with G substituted for F.

In order to express (2.5) in terms of tabulated functions it is convenient to define the following integrals:

$$C_{11}(h, z)\equiv\int_0^h \cos\beta_0 z'(e^{-i\beta_0 R_1}/R_1)dz'; \tag{2.7a}$$

$$C_{12}(h, z, \Delta)\equiv\int_0^h \cos\beta_0 z'(e^{-i\beta_0 R_{12}}/R_{12})dz'; \tag{2.7b}$$

$$S_{11}(h, z)\equiv\int_0^h \sin\beta_0 z'(e^{-i\beta_0 R_1}/R_1)dz'; \tag{2.7c}$$

$$S_{12}(h, z, \Delta)\equiv\int_0^h \sin\beta_0 z'(e^{-i\beta_0 R_{12}}/R_{12})dz'; \tag{2.7d}$$

$$E_{11}(h, z)\equiv\int_0^h (e^{-i\beta_0 R_1}/R_1)dz'; \tag{2.7e}$$

$$E_{12}(h, z, \Delta)\equiv\int_0^h (e^{-i\beta_0 R_{12}}/R_{12})dz'; \tag{2.7f}$$

where

$$R_1\equiv((z'-z)^2+a^2)^{\frac{1}{2}} \tag{2.7g}$$

$$R_2\equiv((z'-z_v)^2+a_v^2)^{\frac{1}{2}}, \tag{2.7h}$$

$$z_v\equiv z\cos\Delta;\quad a_v\equiv(z^2\sin^2\Delta+a^2)^{\frac{1}{2}}. \tag{2.7i}$$

The first four of these functions may be expressed in terms of generalized sine and cosine integrals[2] using

$$U=\beta_0(z'-z);\quad A=\beta_0 a,\quad U_0=\beta_0 z;\quad U_1=\beta_0(h-z), \tag{2.8}$$

and

$$U_v=\beta_0(z'-z_v)=\beta_0(z'-z\cos\Delta);\quad A_v=\beta_0(z^2\sin^2\Delta+a^2)^{\frac{1}{2}}$$
$$U_{0v}=\beta_0 z_v=\beta_0 z\cos\Delta;$$
$$U_{1v}=\beta_0(h-z_v)=\beta_0(h-z\cos\Delta). \tag{2.9}$$

The final formulas for the integrals (2.7a) to (2.7d) are

$$\begin{aligned}C_{11}(h, z)=&\cos U_0[Cc_i(A, U_1)+Cc_i(A, U_0)\\&-jSc(A, U_1)-jSc(A, U_0)]\\&+\sin U_0[-Cs(A, U_1)+Cs(A, U_0)\\&+jSs(A, U_1)-jSs(A, U_0)],\end{aligned} \tag{2.10a}$$

$$\begin{aligned}S_{11}(h, z)=&\sin U_0[Cc_i(A, U_1)+Cc_i(A, U_0)\\&-jSc(A, U_1)-jSc(A, U_0)]\\&+\cos U_0[Cs(A, U_1)-Cs(A, U_0)\\&-jSs(A, U_1)+jSs(A, U_0)],\end{aligned} \tag{2.10b}$$

$$\begin{aligned}C_{12}(h, z, \Delta)=&\cos U_{0v}[Cc_i(A_v, U_{1v})+Cc_i(A_v, U_{0v})\\&-jSc(A_v, U_{1v})-jSc(A_v, U_{0v})]\\&+\sin U_{0v}[-Cs(A_v, U_{1v})+Cs(A_v, U_{0v})\\&+jSs(A_v, U_{1v})-jSs(A_v, U_{0v})],\end{aligned} \tag{2.10c}$$

$$\begin{aligned}S_{12}(h, z, \Delta)=&\sin U_{0v}[Cc_i(A_v, U_{1v})+Cc_i(A_v, U_{0v})\\&-jSc(A_v, U_{1v})-jSc(A_v, U_{0v})]\\&+\cos U_{0v}[Cs(A_v, U_{1v})-Cs(A_v, U_{0v})\\&-jSs(A_v, U_{1v})+jSs(A_v, U_{0v})],\end{aligned} \tag{2.10d}$$

where

$$C_i(A, U)\equiv\sinh^{-1}(U/A)-C(A, U) \tag{2.11}$$

$$Cc_i(A, U)\equiv\sinh^{-1}(U/A)-C(A, U)-Cc(A, U), \tag{2.12}$$

The functions $E_{11}(h, z)$ and $E_{12}(h, z, \Delta)$ defined in (2.7e, f) may be expressed in terms of generalized sine and cosine integrals following the procedure used for the cylindrical antenna.[1] The results are

$$\begin{aligned}E_{11}(h, z)=&C_i(A, U_1)+C_i(A, U_0)\\&-jS(A, U_1)-jS(A, U_0),\end{aligned} \tag{2.13a}$$

$$\begin{aligned}E_{12}(h, z, \Delta)=&C_i(A_v, U_{1v})+C_i(A_v, U_{0v})\\&-jS(A_v, U_{1v})-jS(A_v, U_{0v}).\end{aligned} \tag{2.13b}$$

If the functions defined in (2.5a–d) are expressed in terms of the integrals in (2.7a–f) as expanded in (2.10a–d) and (2.13a, b), they have the following form:

$$\begin{aligned}F_{1v}^{\delta}(z)=&F_{0z}\Psi_v-[C_{11}(h, z)-C_{11}(\delta, z)]\\&+\cos\beta_0 h[E_{11}(h, z)-E_{11}(h, \delta)]\\&+\cos\Delta\{[C_{12}(h, z, \Delta)-C_{12}(\delta, z, \Delta)]\\&-\cos\beta_0 h[E_{12}(h, z, \Delta)-E_{12}(\delta, z, \Delta)]\}\\&+J_v(F_{0z}, z).\end{aligned} \tag{2.14a}$$

$$\begin{aligned}F_{1v}^{\delta}(h)=&-[C_{11}(h, h)-C_{11}(\delta, h)]\\&+\cos\beta_0 h[E_{11}(h, h)-E_{11}(\delta, h)]\\&+\cos\Delta\{[C_{12}(h, h, \Delta)-C_{12}(\delta, h, \Delta)]\\&-\cos\beta_0 h[E_{12}(h, h, \Delta)-E_{12}(\delta, h, \Delta)]\}\\&+J_v(F_{0z}, h).\end{aligned} \tag{2.14b}$$

$$\begin{aligned}G_{1v}^{\delta}(z)=&G_{0z}\Psi_v-[S_{11}(h, z)-S_{11}(\delta, z)]\\&+\sin\beta_0 h[E_{11}(h, z)-E_{11}(h, \delta)]\\&+\cos\Delta\{[S_{12}(h, z, \Delta)-S_{12}(\delta, z, \Delta)]\\&-\sin\beta_0 h[E_{12}(h, z, \Delta)-E_{12}(\delta, z, \Delta)]\}\\&+J_v(G_{0z}, z).\end{aligned} \tag{2.14c}$$

$$\begin{aligned}G_{1v}^{\delta}(h)=&-[S_{11}(h, h)-S_{11}(\delta, h)]\\&+\sin\beta_0 h[E_{11}(h, h)-E_{11}(\delta, h)]\\&+\cos\Delta\{[S_{12}(h, h, \Delta)-S_{12}(\delta, h, \Delta)]\\&-\sin\beta_0 h[E_{12}(h, h, \Delta)-E_{12}(\delta, h, \Delta)]\}\\&+J_v(G_{0z}, h).\end{aligned} \tag{2.14d}$$

If the gap δ is zero, all terms involving δ vanish.

[2] *Tables of Generalized Sine and Cosine-Integral Functions* (Harvard University Press, Cambridge, Massachusetts, 1949), Vols. I and II.

The terms in J_v arising from the particular integral have not been expressed in terms of tabulated functions. However, they may be evaluated by numerical methods.

The expansion parameter Ψ_v defined in (2.1) may be determined readily. Thus

$$\Psi_v(z)=\sin\beta_0 h\int_\delta^h \cos\beta_0 z' K_v(z,z')dz' - \cos\beta_0 h\int_\delta^h \sin\beta_0 z' K_v(z,z')dz'. \quad (2.15)$$

With the definitions in (2.7), the following formula is obtained:

$$\Psi_v(z)=\sin\beta_0 h\{[C_{11}(h,z)-C_{11}(\delta,z)] -\cos\Delta[C_{12}(h,z,\Delta)-C_{12}(\delta,z,\Delta)]\} -\cos\beta_0 h\{[S_{11}(h,z)-S_{11}(\delta,z)] -\cos\Delta[S_{12}(h,z,\Delta)-S_{12}(\delta,z,\Delta)]\}. \quad (2.16)$$

Since only tabulated functions are involved, $\Psi_v(z)$ may be evaluated and plotted to determine the range over which it is sensibly constant at the value $\Psi_v=|\Psi_v(z_r)|$. It is to be expected that an appropriate choice of z_r is $h-\lambda_0/4$ when $\beta_0 h>\pi/2$. On the other hand, the value $z_r=0$ with $\delta=0$, which was appropriate with collinear halves ($\Delta=\pi$), cannot be a good choice. This follows from the fact that A_z decreases rapidly near $z=\delta$. This decrease is not eliminated by letting $\delta=0$ except when $\Delta\doteq\pi$. It follows that for $\beta_0 h\leqq\pi/2$, z_r must not be 0 or δ but somewhere not too close to $z=0$ or $z=h$. A reasonable choice appears to be $z_r=h/2$. That is,

$$\Psi_v=|\Psi_v(z_r)|\begin{cases} z_r=h/2, & \beta_0 h\leqq\pi/2 \\ z_r=h-\lambda_0/4, & \beta_0 h>\pi/2 \end{cases}. \quad (2.17)$$

In evaluating Ψ_v, no significant error is involved if δ is set equal to zero. Application of (2.17) in (2.16) gives, for example, with $\beta_0 h=\pi/2$,

$$\beta_0 h=\pi/2: \quad \Psi_v=|C_{11}(h,\tfrac{1}{2}h)-\cos\Delta C_{12}(h,\tfrac{1}{2}h,\Delta)| \quad (2.18)$$

$$\beta_0 h=\pi/2,\ \Delta=\pi: \quad \Psi_v=|C_{11}(h,h)+C_{12}(h,\tfrac{1}{2}h,\pi)|=C_a(h,\tfrac{1}{2}h), \quad (2.19)$$

where $C_a(h,z)$ is defined in Eq. (11) of the appendix in footnote reference 1. Since $\Psi_v(z)$ is essentially constant over the length of the antenna with collinear halves ($\Delta=\pi$) and zero gap ($\delta=0$), except quite near the ends, it follows that

$$\Psi_v=C_a(h,\tfrac{1}{2}h)\doteq C_a(h,0)=\Psi_{K1}, \quad (2.20)$$

where Ψ_{K1} is the expansion parameter for the straight cylindrical antenna.[1]

General formulas for the current in a V-antenna have been derived. Numerical calculations have not been made.

3. IMPEDANCE OF V-ANTENNA

The general formula for the impedance of the V-antenna is obtained from (2.3) by forming the ratio $V_\delta/I_{1z}(\delta)$. Thus,

$$Z_{v\delta}\equiv V_\delta/I_{1z}(\delta) =\frac{-j\zeta_0\Psi_v}{2\pi}\left\{\frac{\cos\beta_0(h-\delta)+(A_{v1}^\delta/\Psi_v)+\cdots}{\sin\beta_0(h-\delta)+(B_{v1}^\delta/\Psi_v)+\cdots}\right\}, \quad (3.1)$$

where

$$B_{v1}^\delta\equiv M_{v1}^\delta(\delta). \quad (3.2)$$

Numerical evaluation of $Z_{v\delta}$ from (3.1) has not been carried out.

The most important and the simplest special V-antenna is with $\beta_0 h=\pi/2$. For simplicity, also let $\delta=0$, since terminal-zone and gap effects, if significant, can be included in a lumped corrective network if required. The impedance is as follows:

$$Z_{v0}=\frac{-j\zeta_0}{2\pi}\left[\frac{A_{v1}}{1+B_{v1}/\Psi_v}\right] =\frac{-j\zeta_0}{2\pi}\left[\frac{F_{v1}(h)}{1+(1/\Psi_v)[F_{v1}(0)+G_{v1}(h)]}\right], \quad (3.3)$$

where, from (2.5) with (2.7),

$$F_{v1}(0)=\Psi_v-C_{11}(h,0)+\cos\Delta C_{12}(h,0,\Delta), \quad (3.4a)$$

$$F_{v1}(h)=-C_{11}(h,h)+\cos\Delta C_{12}(h,h,\Delta)+J_v(F_{0z},h), \quad (3.4b)$$

$$G_{v1}(h)=-S_{11}(h,h)+\cos\Delta S_{12}(h,h,\Delta)+J_v(G_{0z},h). \quad (3.4c)$$

Substitution of (3.4a–c) in (3.3) leaves the following simple expression for antennas which are sufficiently thin to make Ψ_v quite large:

$$\beta_0 h=\pi/2: \quad Z_{v0}\doteq -j\zeta_0 F_{v1}(h)/2\pi \doteq[j\zeta_0/2\pi][C_{11}(h,h)-\cos\Delta C_{12}(h,h,\Delta) -J_v(\pi/2)]. \quad (3.5)$$

The three functions involved in (3.5) are obtained from the general formulas (2.10a), (2.10c), and (1.33) by setting $z=h$, $\beta_0 h=\pi/2$, and $\delta=0$. Specifically, these functions are as follows:

$$C_{11}(h,h)=[Cs(A,\pi/2)-jSs(A,\pi/2)], \quad (3.6a)$$

$$C_{12}(h,h,\Delta)=\cos U_{0v}[Cc_i(A_v,U_{1v})+Cc_i(A_v,U_{0v}) -jSc(A_vU_{1v})-jSc(A_v,U_{0v})] +\sin U_{0v}[-Cs(A_v,U_{1v})+Cs(A_v,U_{0v}) +jSs(A_v,U_{1v})-jSs(A_v,U_{0v})], \quad (3.6b)$$

where, with $\beta_0 h=\pi/2$, the following definitions apply:

$$U_{0v}=\pi\cos\Delta/2; \quad U_{1v}=\pi(1-\cos\Delta)/2;$$
$$A_v=((\pi\sin\Delta/2)^2+A^2)^{\frac{1}{2}}=\pi\sin\Delta/2. \quad (3.7)$$

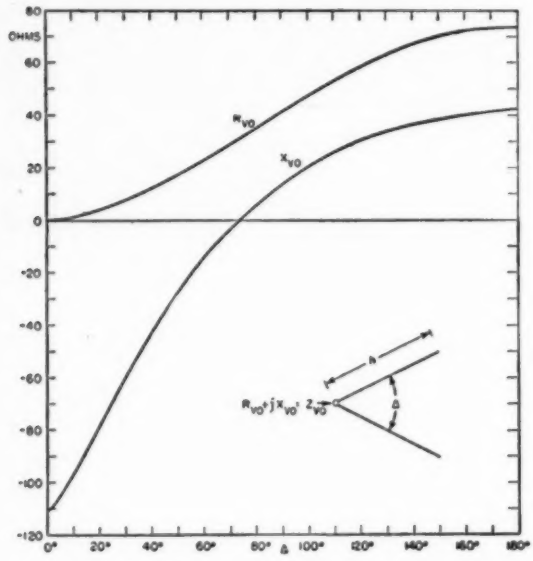


FIG. 2. Zeroth-order impedance of V-antenna; $\beta_0 h=\pi/2$; $\Omega=\infty$.

The approximate formula for A_v on the right in (3.7) is valid when $(\pi \sin\Delta/2)^2 \gg A^2$ or $\sin^2\Delta \gg 16a^2/\lambda_0^2$.

The function $J_v(\pi/2)$ in (3.5) is $J_v(F_{0z'}, h)$ with $\beta_0 h=\pi/2$. With $z=h$, $\beta_0 h=\pi/2$, and $\delta=0$, it is given by

$$J_v(\pi/2)=-\sin^2\Delta\int_0^{\pi/2}\sin u\,du\int_0^{\pi/2}K_{12}(u,u')u'\cos u'\,du', \tag{3.8a}$$

where

$$K_{12}(u,u')=-\frac{e^{-jv}}{v^3}(1+jv), \tag{3.8b}$$

$$v=\beta_0 R_{12}=(u^2-2uu'\cos\Delta+u'^2+A^2)^{\frac{1}{2}}, \tag{3.8c}$$

$$u=\beta_0 w;\quad u'=\beta_0 w';\quad A=\beta_0 a. \tag{3.8d}$$

The term in A^2 is negligible in (3.8c) except when both u and u' are very small.

The exact solution of (3.8a) has not been achieved in closed form except when $\Delta=\pi$ and when Δ is very small. In these two cases, the following values are correct:

$$\Delta=\pi:\quad J_v(\pi/2)=0, \tag{3.9}$$

$A^2 \ll (\pi \sin\Delta/2)^2$; $\Delta^2 \ll 12$:

$$J_v(\pi/2)=(1-\tfrac{1}{2}\Delta^2)\mathrm{Si}\pi=(1-\tfrac{1}{2}\Delta^2)(1.85). \tag{3.10}$$

The value (3.9) follows directly; the value (3.10) may be obtained by direct integration with appropriate approximations.[3] In order to obtain values of $J_v(\pi/2)$ for Δ between π and the upper limit of formula (3.10), (3.8a) was evaluated graphically for $\Delta=\pi/3$, $\pi/2$, $2\pi/3$. In Fig. 2 curves of $Z_{v0}=R_{v0}+jX_{v0}$ as functions of Δ are shown as computed from (3.5). Note that this is a zeroth-order solution ($\Omega=\infty$) for a V-antenna driven from a discontinuity in scalar potential at the apex. With $\Delta=\pi$ it gives the familiar value $Z_v=73.1+j42.5$ for the thin cylindrical antenna. It is readily verified that the value $Z_v=0-j111$ agrees with the value obtained from the theory of nonuniform transmission lines for a uniformly tapered, perfectly conducting line of electrical length $\beta_0 h=\pi/2$.[4]

In practice the V-antenna is not driven by a discontinuity in scalar potential at the apex, but by a potential difference maintained at the junction of a transmission line and the V-antenna. Since the V-antenna is a symmetrical structure, it must be driven from a *balanced* transmission line. Three methods of connecting a V-antenna to a two-wire line are shown in Fig. 3. In general, the apparent terminal impedances for the lines in Figs. 3(a) and 3(b), differ significantly from each other and from the impedance of the idealized V-antenna driven by a slice generator. These differences

[3] In the range of small values of Δ, v assumes correspondingly small values when $|u'-u|$ is small. Hence, the significant contributions to the integrand in (3.8a) are when v is small and $K_{12}(u,u')$ large. The appropriate approximations are

$$\Delta^2 \ll 12: \cos\Delta \doteq 1-\tfrac{1}{2}\Delta^2;\quad v\doteq(u^2-2uu'(1-\tfrac{1}{2}\Delta^2)+u'^2)^{\frac{1}{2}} \tag{3.10a}$$

$$v^2 \ll 12: K_{12}(u,u')\doteq -1/v^3. \tag{3.10b}$$

It is assumed in (3.10) that contributions from A^2 are negligible. Since the principal contributions to the integral occur when u' is near u, no serious error is made if u' is replaced by u in the slowly varying factor $\cos u'$. With (3.10a) and (3.10b) and $\cos u'=\cos u$,

$$J_v(\pi/2)\doteq\Delta^2\int_0^{\pi/2}\sin u\cos u\,du\int^{\pi/2}u'\,du'/v^3. \tag{3.10c}$$

The u' integral integrates directly using standard formulas (e.g., Pierce, 170). Retaining terms in Δ^2 the result is

$$\int^{\pi/2}u'\,du'/v^3\doteq-\tfrac{1}{4}\pi[1/u(\tfrac{1}{2}\pi-u)]+2/\Delta^2 u. \tag{3.10d}$$

Substitutions of (3.10d) in 3.10c) and integration give

$$J_v(\pi/2)=(1-\tfrac{1}{2}\Delta^2)\mathrm{Si}\pi=(1-\tfrac{1}{2}\Delta^2)(1.85). \tag{3.10e}$$

[4] It is shown by S. A. Schelkunoff (*Electromagnetic Waves* (D. Van Nostrand Company, Inc., New York, 1943), pp. 239 and 293) that the resonant electrical length of a tapered line open at $z=h$ and closed at $z=0$ is

$$\beta_0 h_r=\tfrac{1}{2}\pi(1+\delta), \tag{3.10f}$$

where,

$$\delta=-\frac{1}{\pi R_{c0}}\int^{\pi}R_c(y)\cos y\,dy;\quad y=2\beta_0 z. \tag{3.10g}$$

In (3.10g)

$$R_c(y)=120\ln(2z\sin\tfrac{1}{2}\Delta/a)=120\ln(y\sin\tfrac{1}{2}\Delta/A) \tag{3.10h}$$

is the variable characteristic impedance of the lossless tapered line, and R_{c0} is the average characteristic impedance. Integration by parts gives directly:

$$\delta=120\mathrm{Si}\pi/\pi R_{c0}. \tag{3.10i}$$

The input reactance of a section of lossless uniform transmission line of length s is

$$X_{in}=-R_c\cot\beta_0 s. \tag{3.10j}$$

Since the resonant electrical length of the tapered line is greater than $\pi/2$ by $\delta\pi/2$, the input reactance of a section of length $\pi/2$ is given approximately by the input reactance of a section of uniform line with characteristic impedance R_{c0} and of electrical length $\delta\pi/2$ below resonance. For the uniform-line resonance is at $\pi/2$, so that the length $(\pi/2)(1-\delta)$ is related to $\pi/2$ approximately as, for the tapered line, $\pi/2$ is related to $(\pi/2)(1+\delta)$. Hence, with $\delta^2 \ll 1$,

$$X_{in}\doteq-R_{c0}\cot\tfrac{1}{2}\pi(1-\delta)=-R_{c0}\tan\tfrac{1}{2}\pi\delta\doteq-R_c\pi\delta/2. \tag{3.10k}$$

Using the above value of δ,

$$X_{in}\doteq-60\mathrm{Si}\pi\text{ ohms}=-111\text{ ohms}. \tag{3.10l}$$

This agrees exactly with the value obtained from (3.5) with Δ very small.

are due to transmission-line end effects and to coupling between the V-antenna and the transmission line. Only when b/λ_0 is extremely small (0.001 or less) do the apparent terminal impedances for the lines in Figs. 3(a) and 3(b) approximate each other and the theoretical value for the slice generator. Note that this last is the impedance as b approaches zero. In the circuit of Fig. 3(c), on the other hand, there are no antenna-line coupling effects, and transmission-line end effects may be compensated by adjusting the stub so that a voltage maximum is maintained across the terminals of the antenna when this is disconnected. If this is done and the inductive reactance of the two half-bridges joining the line wires to the antenna is small, the apparent terminal impedance loading the line in Fig. 3(c) should be approximated closely by the theoretical value for a V-antenna driven from a slice generator.

The accurate calculation of the apparent terminal impedance as a function of Δ in either of the circuits in Figs. 3(a, b) is difficult if b/λ_0 is not extremely small. However, it is possible to derive the general nature and behavior of the impedance in a semiquantitative manner by modification of the zeroth-order solution of the ideal apex-driven V-antenna. This involves determining the equivalent lumped elements of terminal-zone networks for the practical circuits of Figs. 3(a, b), and combining these with the theoretical impedance for the ideal isolated V-antenna shown in Fig. 1. Note that both h and δ are functions of Δ, whereas $h'=h-\delta$ is constant. As Δ approaches zero, δ becomes infinite, since when $\Delta \doteq 0$, the antenna structure ceases to be a tapered line and becomes a section of two-wire line. Although a general formula for the impedance of the V-antenna has been derived with both δ and h parameters, the actual evaluation in a quantitative sense is difficult since δ does not remain small. An instructive and valuable estimate of the variation of the impedance with the enclosed angle Δ may be obtained from the zeroth-order solution for the idealized V, the terminal-zone networks for Figs. 3(a, b) and the transmission-line impedance at $\Delta=0$. This is true in particular when $\beta_0 h=\pi/2$. Let the two circuits illustrated in Figs. 3(a, b) be analyzed successively. Since the circuit of Fig. 3(b) is slightly simpler, it is considered first.

Antenna Perpendicular to Plane of Two-Wire Line

In the circuit of Fig. 3(b) there is no inductive coupling between antenna and line. Therefore, the series inductive element L_T of the terminal-zone network is[5]

$$L_T \doteq -(b-a)/2\pi v_0. \tag{3.11}$$

The evaluation of the shunt capacitive element C_T differs from C_{Te}[5] for $\Delta=\pi$, only in the formulas for

[5] R. King, "Theory of antennas driven from a two-wire line," Cruft Laboratory Technical Report No. 41, and J. Appl. Phys. 20, 832–850 (1949).

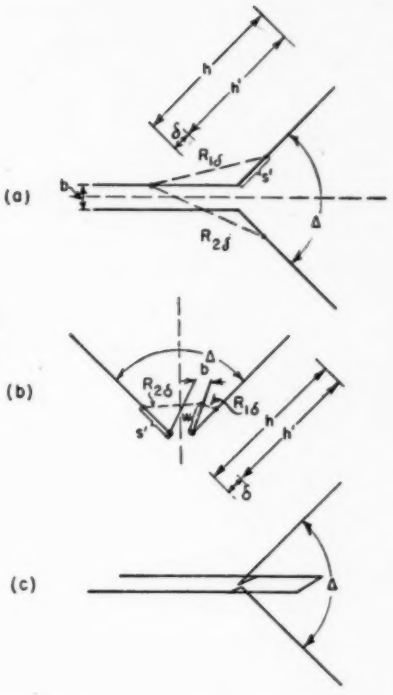


FIG. 3. Three methods of driving a V-antenna from a two-wire line.

R_{1T} and R_{2T}. The following expressions apply:

$$R_{1T}=(s'^2+w^2+a^2)^{\frac{1}{2}};$$

$$R_{2T} \doteq (s'^2+2s'b\sin\tfrac{1}{2}\Delta+w^2+b^2+a^2)^{\frac{1}{2}}. \tag{3.12}$$

As shown in Fig. 3(b), w is the distance along the line to the point of calculation of the potential, s' is the distance along the antenna to the element of charge $q'ds'$. Both these distances are measured from the junction of the antenna and the line. The line spacing is b, the radius of all conductors is a. The following inequality is assumed: $a^2 \ll b^2$. C_T is given by

$$C_T=\int_0^d [c(w)-c_0]dw, \tag{3.13}$$

where $d \doteq 10b$ and where

$$[c(w)-c_0]/c_0$$

$$=-\left[\frac{\sinh^{-1}(w/a)-\sinh^{-1}(w/b)+k_q \ln B-\ln(b/a)}{\sinh^{-1}(w/a)-\sinh^{-1}(w/b)+k_q \ln B+\ln(b/a)}\right]. \tag{3.14}$$

The charge-ratio factor $k_q \doteq 1$ when $\beta_0 h$ is near π. When $\beta_0 h$ is near $\pi/2$, the effect of C_T in parallel is relatively unimportant so that C_T may be omitted.

The accurate determination of the apparent impedance terminating the line in the circuit of Fig. 3(b) is involved. The following steps are required: (1) The evaluation of the theoretical isolated impedance $Z_{v\delta}$ from (1) for the appropriate values of h and Δ. Note that $\delta=(b/2)\csc(\Delta/2)$, so that δ is large if Δ is small. (2) Combination of $Z_{v\delta}$ with the terminal-zone network

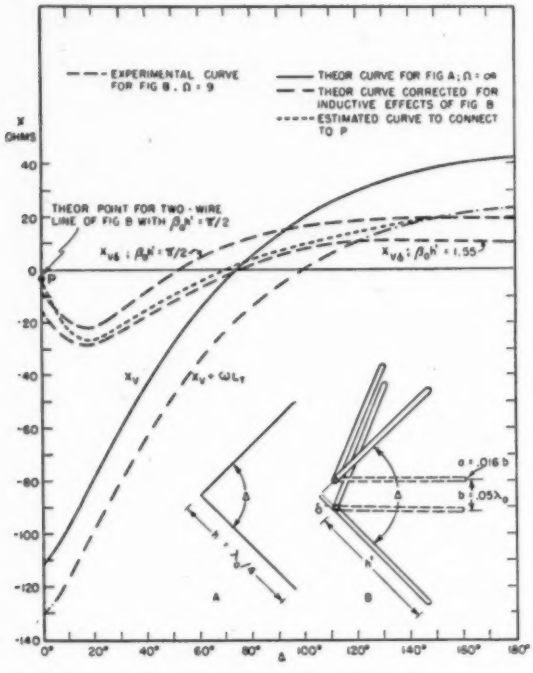

EXPERIMENTAL CURVE FOR FIG B, Ω = 9
THEOR CURVE FOR FIG A; Ω = 9
THEOR CURVE CORRECTED FOR INDUCTIVE EFFECTS OF FIG B
ESTIMATED CURVE TO CONNECT TO P
X OHMS
THEOR POINT FOR TWO-WIRE LINE OF FIG B WITH $\beta_0 h' = \pi/2$
$X_{v\delta}$; $\beta_0 h' = \pi/2$
$X_{v\delta}$; $\beta_0 h' = 1.55$
P
X_v
$X_v + \omega L_T$
a = .016 b
b = .05λ0
Δ
A
B
0° 20° 40° 60° 80° Δ 100° 120° 140° 160° 180°

FIG. 4. Reactance of V-antenna.

consisting of L_T in series and C_T in parallel with $Z_{v\delta}$. Since numerical values of $Z_{v\delta}$ are available only from a zeroth-order solution with $\delta=0$ and $\beta_0 h=\pi/2$, an accurate quantitative determination of the apparent terminal impedance in Fig. 3(b) is not possible at present. Nevertheless, a valuable qualitative picture of the apparent impedance as a function of Δ may be obtained for $\beta_0 h'=\beta_0(h-\delta)=\pi/2$ by applying the terminal-zone correction to the theoretical zeroth-order impedance for $\delta=0$. As an example, consider a rather thick antenna ($\Omega=2\ln(2h/a)\doteq 9$) driven from a two-wire line that is quite widely spaced so that $b=0.05\lambda_0$, $a=0.016b$. With $f=750$ mc/sec,

$$\omega L_T \doteq -b/2\pi\nu_0 \doteq -18.8 \text{ ohms.} \qquad (3.15a)$$

In Fig. 4 is shown the zeroth-order curve for X_v for $\delta=0$, $\beta_0 h=\pi/2$ as taken from Fig. 2, together with a curve of $X_v+\omega L_T$. Since for $\Delta=\pi$ at $\delta=0$ the zeroth-order reactance, 42.5 ohms, is a fair approximation of the reactance even of quite thick antennas, and since the principal effect of a finite but small gap 2δ on the theoretical terminal impedance is taken into account by using $h'=h-\delta$ in place of h, it follows that $X_v+\omega L_T$ in Fig. 4 should be a rough approximation of $X_{v\delta}$ when Δ is sufficiently near π so that $\beta_0\delta=(\beta_0 b/2)\csc(\Delta/2)$ is small compared with unity. With $\beta_0 b$ as large as 0.314, the acceptable range is limited to $\Delta\leqq 120°$. Evidently $X_v+\omega L_T$ is a reasonable approximation of the apparent terminal reactance over a much greater range of Δ when smaller values of $\beta_0 b$ and of L_T are involved. However, $X_v+\omega L_T$ is in no case a good approximation for small values of Δ, since the assumption $\delta=0$ implied in X_v is then not a good one. For $\Delta\doteq 0$, the impedance of the antenna reduces to the impedance of a section of two-wire line of electrical length $\beta_0 h=\pi/2$. Conventional transmission-line theory gives zero impedance for such a section if the conductors are treated as perfect. However, if account is taken of the capacitive end effect, a lumped positive capacitance C_T is required across the open end of the section. This capacitance is given by (3.13) with (3.14) provided k_q is set equal to zero. Graphical integration with $b/a=6$ gives $C_T=0.163\times 10^{-12}$ farad. The reactance looking into a quarter-wave section of lossless line terminated in C_T is

$$X_{in}=\omega C_T R_c^2=35 \text{ ohms,} \qquad (3.15b)$$

where $R_c=215.4$ ohms. This is the input reactance at an electrical distance $\beta_0 w=\pi/2$ from the end of a straight transmission line. Since there is a right-angle bend at this value of $\beta_0 w$, the apparent impedance involves a terminal-zone network. At a low voltage point the capacitive effect is small, but the inductive effect is great. In order to correct uniform line theory on each side of the bend, series inductances L_T are required. Thus, the reactance terminating the line is X_{in} in (3.15b) in series with $2L_T$ where L_T is the same as (3.15a). Thus,

$$X_{sa}=X_{in}+2\omega L_T=35-37.6=-2.6 \text{ ohms.} \qquad (3.16)$$

This point is indicated by P in Fig. 4. Evidently, between $\Delta=120°$ and $\Delta=0°$, the curve $X_v+\omega L_T$ must bend up so that it ends at $X_{sa}=-2.6$, the value for the two-wire line, instead of at $X=-111$, the value for a tapered line. It is to be expected that the true curve tends to follow $X_v+\omega L_T$ but ultimately bends up to

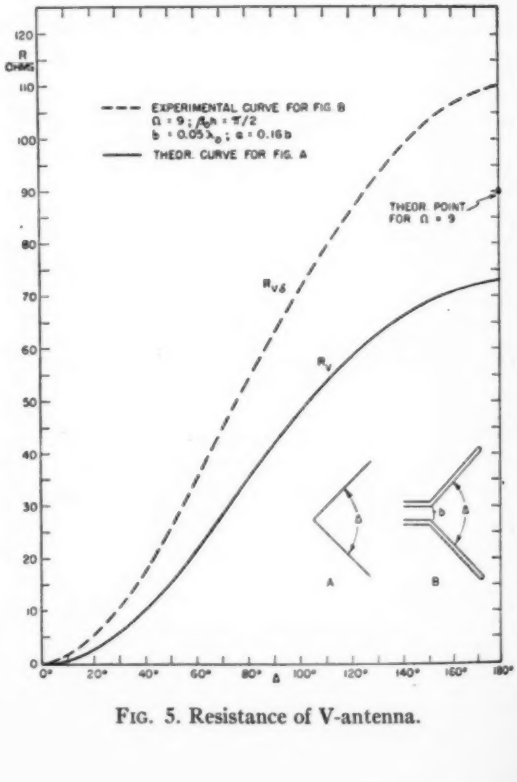


FIG. 5. Resistance of V-antenna.

the point P. For a relatively thick antenna a curve such as the dotted one in Fig. 4 appears reasonable. It is seen to approximate the experimentally determined[6] curve marked $\beta_0 h' = \pi/2$. An experimental curve for a slightly smaller value of h' is marked $\beta_0 h' = 1.55$. The presently used rough method of determining the approximate reactance curve of a thick V-antenna driven from a two-wire line with large spacing provides a clear picture of the several factors that determine the shape of the experimental curve and the striking departure from the simple curve X_v of the thin V-antenna driven from a two-wire line with negligible spacing b. In addition, the great significance of terminal-zone effects in determining the reactance of the V-antenna is clarified.

The terminal-zone effects do not alter R_v significantly. It is seen in Fig. 5 that the zeroth-order theo-

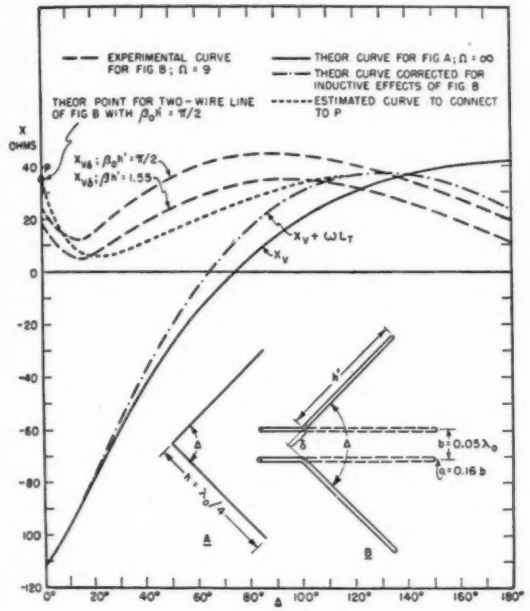

EXPERIMENTAL CURVE FOR FIG. B; Ω = 9
THEOR CURVE FOR FIG. A; Ω = ∞
THEOR CURVE CORRECTED FOR INDUCTIVE EFFECTS OF FIG. B
THEOR POINT FOR TWO-WIRE LINE OF FIG. B WITH $\beta_0 h' = \pi/2$
ESTIMATED CURVE TO CONNECT TO P
X OHMS
$X_{vδ}; \beta_0 h' = \pi/2$
$X_{vδ}; \beta_0 h' = 1.55$
$X_v + \omega L_T$
X_v
$b = 0.05\lambda_0$
$a = 0.16\,b$
$h = \lambda_0/4$
A
B
Δ

FIG. 6. Reactance of V-antenna.

retical curve R_v for the extremely thin antenna with $\delta = 0$ and $\beta_0 h = \pi/2$ agrees in shape with the experimental curve $R_{v\delta}$ for a V-antenna with $\Omega = 9$, $\beta_0 h' = \beta_0(h-\delta) = \pi/2$ driven from a two-wire line with $b/\lambda_0 = 0.05$, $a = 0.16b$. Numerical differences for $\Delta < \pi$ are consistent with the difference at $\Delta = \pi$ where the theoretical value is $R_v = R_0 = 73.1$ ohms for a very thin antenna.

Antenna in the Plane of a Two-Wire Line

An estimate of the apparent impedance terminating the line in the circuit of Fig. 3(a) is more involved than for the circuit of Fig. 3(b), since inductive as well as capacitive coupling existing between the antenna and the transmission line. The lumped elements L_T and

[6] Experimental data are those of D. Angelakos.

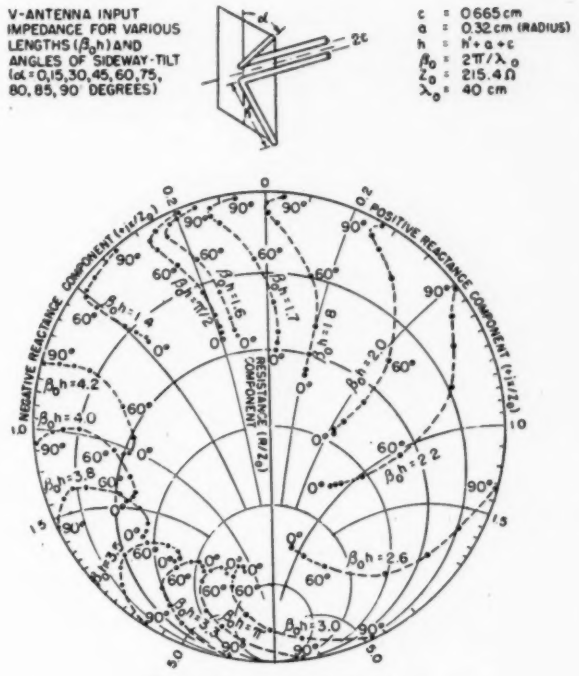

V-ANTENNA INPUT IMPEDANCE FOR VARIOUS LENGTHS ($\beta_0 h$) AND ANGLES OF SIDEWAY-TILT (α = 0, 15, 30, 45, 60, 75, 80, 85, 90 DEGREES)
c = 0.665 cm
a = 0.32 cm (RADIUS)
h = h' + a + c
$\beta_0 = 2\pi/\lambda_0$
Z_0 = 215.4 Ω
λ_0 = 40 cm
NEGATIVE REACTANCE COMPONENT ($\pm jX/Z_0$)
POSITIVE REACTANCE COMPONENT ($+jX/Z_0$)
RESISTANCE COMPONENT (R/Z_0)

FIG. 7. Sideway-tilt V-antenna.

C_T for the corrective-terminal-zone network are determined by the same general method. Thus, the lumped shunt capacitance C_T is given by (3.13) and (3.14) but with[7]

$$\ln B = k_0 T'(w) = \int_0^\infty [(1/R_{1T}) - 1/R_{2T}]\,ds'$$

$$= \ln\left[\frac{w\cos\tfrac{1}{2}\Delta + b\sin\tfrac{1}{2}\Delta + (w^2+b^2)^{\frac{1}{2}}}{w\cos\tfrac{1}{2}\Delta + (w^2+a^2)^{\frac{1}{2}}}\right], \quad (3.17a)$$

where

$$R_{1T} = (s'^2 + 2s'w\cos\tfrac{1}{2}\Delta + w^2 + a^2)^{\frac{1}{2}};$$

$$R_{2T} = (s'^2 + 2s'(w\cos\tfrac{1}{2}\Delta + b\sin\tfrac{1}{2}\Delta) + w^2 + b^2)^{\frac{1}{2}} \quad (3.17b)$$

in place of (3.13).

The series inductance L_T involves both end effect and coupling. It is determined from the inductance per unit length of the two-wire line with the V-antenna attached, i.e., from

$$L_T = \int_0^d [l^e(w) - l_0{}^e]\,dw$$

$$= \frac{1}{2\pi v_0}\left\{\cos\tfrac{1}{2}\Delta \int_0^d \ln\left[\frac{w\cos\tfrac{1}{2}\Delta + b\sin\tfrac{1}{2}\Delta + (w^2+b^2)^{\frac{1}{2}}}{w\cos\tfrac{1}{2}\Delta + (w^2+a^2)^{\frac{1}{2}}}\right]dw \right.$$

$$\left. - \int_0^d \ln\left[\frac{w + (w^2+b^2)^{\frac{1}{2}}}{w + (w^2+a^2)^{\frac{1}{2}}}\right]dw\right\}, \quad (3.18)$$

where $d \geqq 10b$.

[7] R. King and K. Tomiyasu, "Terminal impedance and generalized two-wire line theory," Cruft Laboratory Technical Report No. 74, and Proc. Inst. Radio Engrs. **37**, 1134–1139 (1949).

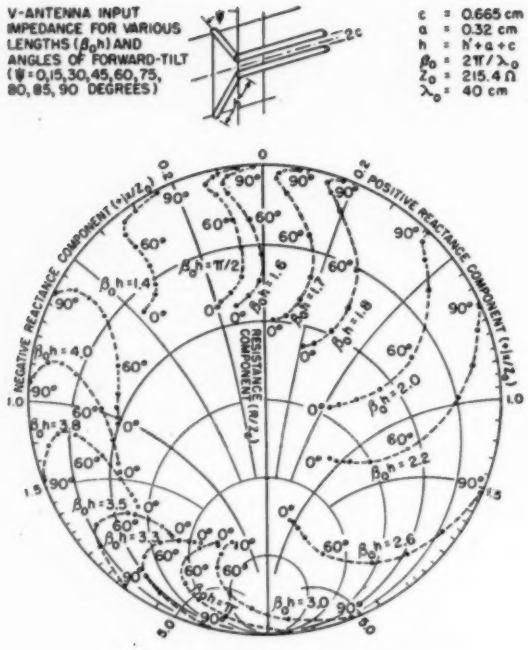

V-ANTENNA INPUT IMPEDANCE FOR VARIOUS LENGTHS ($\beta_0 h$) AND ANGLES OF FORWARD-TILT (Ψ = 0, 15, 30, 45, 60, 75, 80, 85, 90 DEGREES)
c = 0.665 cm
a = 0.32 cm
h = h'+a+c
β_0 = 2π/λ_0
Z_0 = 215.4 Ω
λ_0 = 40 cm
POSITIVE REACTANCE COMPONENT (+jX/Z_0)
NEGATIVE REACTANCE COMPONENT (+jX/Z_0)
RESISTANCE COMPONENT (R/Z_0)

FIG. 8. Forward-tilt V-antenna.

The combination $X_v+\omega L_T$ is shown in Fig. 6 for the same antenna and line analyzed previously for the circuit of Fig. 3(b). This curve for the circuit of Fig. 3(a) is seen to differ considerably from the corresponding curve for the circuit of Fig. 3(b) shown in Fig. 4. Notably, the reactance increases rather than decreases as Δ is reduced from π. Since the theoretical curves assume $\delta=0$, $X_v+\omega L_T$ can be expected to be a fair approximation for thick antennas with $\delta<0$ only for Δ sufficiently large. For the same line spacing, $b=0.05\lambda$, the range $120°\leqq\Delta\leqq180°$ is reasonable.

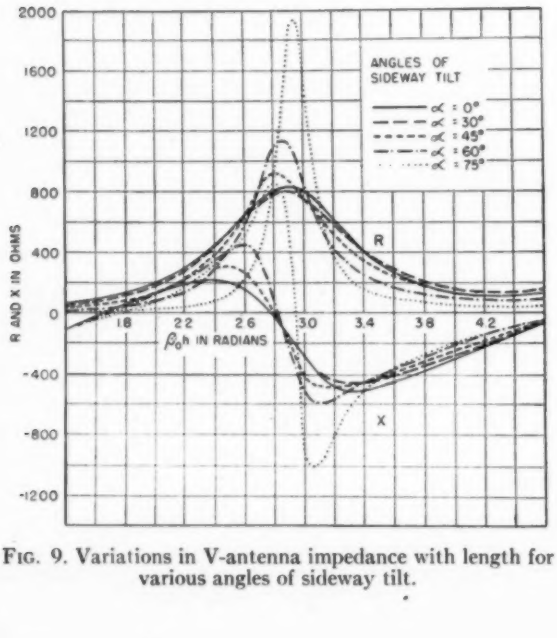


FIG. 9. Variations in V-antenna impedance with length for various angles of sideway tilt.

Just as in the case of Fig. 3(b), the reactance at $\Delta=0$ may be obtained from transmission-line theory appropriately corrected for end effects. In this case there is no right-angle bend at the junction of the line proper and the section corresponding to V-antenna folded together so that $\Delta=0$; therefore, only the terminal-zone effect at the end of the section of length $\pi/2$ is effective. This is the same as before and gives

$$X_{sa}=X_{in}=\omega C_T R_c^2=35 \text{ ohms}, \qquad (3.19)$$

for the particular case under study. The point $X=35$ is labeled P in Fig. 6. As before, the true impedance curve must connect $X_v+\omega L_T$ between $\Delta=120°$ and 180° with this point. An estimated curve of this sort is shown dotted. Experimental curves[6] of the apparent impedance of the V-antenna for $\beta_0 h'=\pi/2$ and 1.55 are marked $R_{v\delta}$ in Fig. 6.

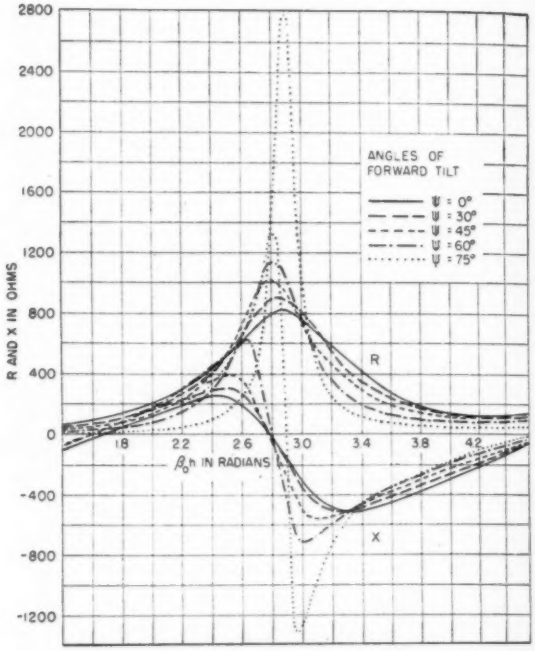


FIG. 10. Variations in V-antenna impedance with length for various angles of forward tilt.

The analysis has made it clear how important terminal-zone effects are when b/λ_0 is not vanishingly small. The rough qualitative study based on the theoretical zeroth-order reactance X_v is adequate to explain the shape of the experimental curve and the reasons for its departure not only from the theoretical zeroth-order curve for the ideal V-antenna but also from the experimental curve for the identical antenna driven from the same line but in the circuit of Fig. 3(b) instead of that of Fig. 3(a). The great importance of the particular circuit connection in determining the apparent reactance of a V-antenna is brought our clearly. The resistance of the V-antenna is essentially the same

in both the circuits in Fig. 3(a) and Fig. 3(b). It is given in Fig. 5.

ACKNOWLEDGMENTS

The author is grateful to Dr. C. T. Tai and Mr. J. E. Storer for valuable discussions.

APPENDIX. EXPERIMENTAL MEASUREMENT OF THE IMPEDANCE OF V-ANTENNAS

Complete experimental curves for the apparent impedance of the V-antenna drives with the antenna in the plane of the line and perpendicular to the plane of the line are in Figs. 7 to 10. Note the striking differences between the measured impedances of the *same* antenna when driven from the same line in the two orientations. The two sets of curves agree only when the two legs are collinear and the two orientations become identical.

The experimental measurements were made by D. Angelakos[8] on the image-plane line described elsewhere.[9] The notation on Figs. 7–10 differs somewhat from that used in the main body of this paper. The electrical leg length $\beta_0 h$ used as parameter is the electrical half-length of the antenna when the legs are collinear. That is, in these figures $\beta_0 h = \beta_0(h'+a+c)$, where $2(a+c)$ is the line spacing b. Note that Δ, the enclosed angle of the V, is represented in the form $(\pi - 2\psi)$ for the antenna in the plane of the line and $(\pi - 2\alpha)$ for the antenna perpendicular to the plane of the line. The angles ψ and α are the angles of tilt of each leg of the V from the collinear position.

[8] D. Angelakos, "Current and charge distributions on antennas and open-wire lines," Ph.D. thesis (Cruft Laboratory, Harvard University, Cambridge, Massachusetts).

[9] P. Conley, "Impedance measurements with open-wire lines" (Cruft Laboratory Technical Report No. 35, March 18, 1948). P. Conley, J. Appl. Phys. **20**, 1022–1026 (1949).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 9 SEPTEMBER, 1951

On the Stress-Function Approaches of Boussinesq and Timpe to the Axisymmetric Problem of Elasticity Theory*

E. Sternberg,† R. A. Eubanks,‡ and M. A. Sadowsky§
Illinois Institute of Technology, Chicago, Illinois

(Received March 14, 1951)

In the first part of this note the stress-function approaches of Boussinesq and Timpe to the rotationally symmetric problem in the classical theory of elasticity are referred to general orthogonal axisymmetric curvilinear co-ordinates. Although the underlying computations are of a routine character, it is felt that the presentation of the final results may serve a useful purpose. In the second part of the note, a connection between the two stress-function approaches is established.

INTRODUCTION

THE torsion-free rotationally symmetric problem in the theory of elasticity has been attacked by means of stress functions in a variety of ways.[1] The general solution of the displacement equations of equilibrium in terms of three harmonic functions, which was apparently initiated by Boussinesq [3],[2] in case of rotational symmetry yields an approach via two harmonic functions. Probably the most widely used approach is that due to Love [1] in terms of a single biharmonic function. Love's solution, moreover, coincides with the general solution of Galerkin [2] in the special instance of rotational symmetry, as was pointed out by H. M. Westergaard.[3] A variant of Love's solution was given by A. and L. Föppl [4]. More recently, Timpe [10] and Weber [11] have introduced further stress-function approaches to the axisymmetric problem.

The relative merits of the general solutions referred to above depend upon the particular axisymmetric problem under consideration. For applications involving the use of noncylindrical curvilinear co-ordinates, the stress functions of Boussinesq and Timpe appear to offer definite advantages over those of Love and Weber, which give rise to comparatively unwieldy expressions for the curvilinear components of stress. In what follows we exhibit explicitly the solutions of Boussinesq and Timpe in general orthogonal axisymmetric co-ordinates.[4]

From the point of view of applications, it furthermore seems useful to examine the connection between the various approaches. The connection between the solutions of Boussinesq and Love was given by Mindlin [9]. The relation between Weber's solution and that of Boussinesq is also immediate, as is pointed out by Butty [2]. In the current note we relate Timpe's approach to that of Boussinesq, and thus to the remaining approaches cited.

* The results presented in this paper were obtained in the course of an investigation conducted under Contract N7onr-32906 with the ONR, Department of the Navy, in Washington, D. C.

† Associate Professor of Mechanics, Illinois Institute of Technology, Chicago, Illinois.

‡ Research Assistant, Illinois Institute of Technology, Chicago, Illinois.

§ Associate Professor of Mathematics, Illinois Institute of Technology, Chicago, Illinois.

[1] See [2] (numbers in brackets refer to the bibliography at the end of the paper), which contains a comprehensive treatment of this subject.

[2] The three-function approach was independently established by Papkovich [6] and subsequently by Neuber [7] who has employed it extensively [8]. A particularly elegant completeness proof for the Bossinesq solution was given by Mindlin [9].

[3] See [13] in which this observation of Westergaard is cited.

[4] Although the curvilinear transforms of the corresponding displacement fields are available elsewhere [2], [10], they will be included here for the sake of convenience.

BOUSSINESQ'S APPROACH

In accordance with the three-function approach of Boussinesq [3], the general solution of the displacement equations of equilibrium in case of rotational symmetry about the z axis, and in the absence of body forces, is representable as the sum of the following two displacement fields which are here referred to the cylindrical co-ordinates (ρ, γ, z):

$$2G[u_\rho, u_\gamma, u_z] = \text{grad}\,\varphi, \tag{1}$$

$$2G[u_\rho, u_\gamma, u_z] = \text{grad}(z\psi) - [0, 0, 4(1-\nu)\psi], \tag{2}$$

where

$$\nabla^2\varphi(\rho, z) = 0, \quad \nabla^2\psi(\rho, z) = 0, \tag{3}$$

and $[u_\rho, u_\gamma, u_z]$ denote the radial, transverse, and axial components of displacement, whereas G and ν designate the shear modulus and Poisson's ratio, respectively. The two solutions generated by the stress functions φ and ψ will henceforth be called the first and second Boussinesq solutions.

We now introduce general axisymmetric curvilinear co-ordinates (α, β, γ) by means of the transformation

$$\left.\begin{aligned} x &= f(\alpha, \beta)\cos\gamma \\ y &= f(\alpha, \beta)\sin\gamma \\ z &= g(\alpha, \beta) \end{aligned}\right\}, \tag{4}$$

which is assumed to conform to the orthogonality condition

$$f_\alpha f_\beta + g_\alpha g_\beta = 0 \tag{5}$$[5]

so that the coordinate surfaces $\alpha = \text{const}$, $\beta = \text{const}$ represent mutually orthogonal families of coaxial surfaces of revolution which are in turn perpendicular to the meridional half-planes $\gamma = \text{const}$. The differential of arc-length is given by

$$ds^2 = \left(\frac{d\alpha}{h_1}\right)^2 + \left(\frac{d\beta}{h_2}\right)^2 + \left(\frac{d\gamma}{h_3}\right)^2, \tag{6}$$

and the metric coefficients become

$$h_1 = (f_\alpha^2 + g_\alpha^2)^{-\frac{1}{2}}, \quad h_2 = (f_\beta^2 + g_\beta^2)^{-\frac{1}{2}}, \quad h_3 = \frac{1}{|f|}. \tag{7}$$

By the usual transformation methods [12], we obtain for the displacement fields of the Boussinesq solutions in curvilinear coordinates,

$$\left.\begin{aligned} u_\alpha &= h_1\varphi_\alpha/2G \\ u_\beta &= h_2\varphi_\beta/2G \end{aligned}\right\}, \tag{8}$$

$$\left.\begin{aligned} u_\alpha &= (h_1/2G)[g\psi_\alpha - (3-4\nu)g_\alpha\psi] \\ u_\beta &= (h_2/2G)[g\psi_\beta - (3-4\nu)g_\beta\psi] \end{aligned}\right\}, \tag{9}$$

and for the associated fields of stress,

$$\left.\begin{aligned} \sigma_\alpha &= h_1^2\varphi_{\alpha\alpha} + h_1\frac{\partial h_1}{\partial\alpha}\varphi_\alpha - \frac{h_2^2}{h_1}\frac{\partial h_1}{\partial\beta}\varphi_\beta \\ \sigma_\beta &= h_2^2\varphi_{\beta\beta} + h_2\frac{\partial h_2}{\partial\beta}\varphi_\beta - \frac{h_1^2}{h_2}\frac{\partial h_2}{\partial\alpha}\varphi_\alpha \\ \sigma_\gamma &= \frac{h_1^2 f_\alpha}{f}\varphi_\alpha + \frac{h_2^2 f_\beta}{f}\varphi_\beta \\ \tau_{\alpha\beta} &= h_1h_2\varphi_{\alpha\beta} + h_2\frac{\partial h_1}{\partial\beta}\varphi_\alpha + h_1\frac{\partial h_2}{\partial\alpha}\varphi_\beta \end{aligned}\right\}, \tag{10}$$

$$\left.\begin{aligned} \sigma_\alpha &= h_1^2 g\psi_{\alpha\alpha} + \left(h_1\frac{\partial h_1}{\partial\alpha}g - 2h_1^2 g_\alpha\right)\psi_\alpha \\ &\quad - \frac{h_2^2}{h_1}\frac{\partial h_1}{\partial\beta}g\psi_\beta + 2\nu(h_1^2 g_\alpha\psi_\alpha - h_2^2 g_\beta\psi_\beta) \\ \sigma_\beta &= h_2^2 g\psi_{\beta\beta} + \left(h_2\frac{\partial h_2}{\partial\beta}g - 2h_2^2 g_\beta\right)\psi_\beta \\ &\quad - \frac{h_1^2}{h_2}\frac{\partial h_2}{\partial\alpha}g\psi_\alpha + 2\nu(h_2^2 g_\beta\psi_\beta - h_1^2 g_\alpha\psi_\alpha) \\ \sigma_\gamma &= \frac{g}{f}(h_1^2 f_\alpha\psi_\alpha + h_2^2 f_\beta\psi_\beta) - 2\nu(h_1^2 g_\alpha\psi_\alpha + h_2^2 g_\beta\psi_\beta) \\ \tau_{\alpha\beta} &= h_1h_2 g\psi_{\alpha\beta} + g\left(h_1\frac{\partial h_2}{\partial\alpha}\psi_\beta + h_2\frac{\partial h_1}{\partial\beta}\psi_\alpha\right) \\ &\quad - (1-2\nu)h_1h_2(g_\alpha\psi_\beta + g_\beta\psi_\alpha). \end{aligned}\right\} \tag{11}$$

In the foregoing equations u_α, u_β and σ_α, σ_β, σ_γ, $\tau_{\alpha\beta}$ are the curvilinear components of displacement and stress, respectively; the displacement u_γ as well as the stresses $\tau_{\gamma\alpha}$, $\tau_{\gamma\beta}$ vanish identically by virtue of the rotational symmetry. Finally, we record the curvilinear transform of Laplace's equation, which must be satisfied by the Boussinesq stress functions. If $\phi(\alpha, \beta)$ is harmonic,

$$\frac{\partial}{\partial\alpha}\left(\frac{fh_1}{h_2}\phi_\alpha\right) + \frac{\partial}{\partial\beta}\left(\frac{fh_2}{h_1}\phi_\beta\right) = 0. \tag{12}$$

TIMPE'S APPROACH

We next consider the stress-function approach due to Timpe [10]. Timpe represented the general axisymmetric integral of the displacement equations of equilibrium as the sum of two displacement fields which, in cylindrical coordinates, appear as

$$2G[u_\rho, u_\gamma, u_z] = \text{grad}\,\Phi, \tag{13}$$

$$2G[u_\rho, u_\gamma, u_z] = \text{grad}\,\Psi - \left[\frac{4(1-\nu)}{\rho}\Psi, 0, 0\right], \tag{14}$$

where

$$\left.\begin{aligned} \nabla^2\Phi(\rho, z) &= 0, \\ \Lambda^2\Psi(\rho, z) \equiv \nabla^2\Psi - \frac{2}{\rho}\frac{\partial\Psi}{\partial\rho} &= 0. \end{aligned}\right\} \tag{15}$$

[5] Subscripts attached to functions which originally bear no subscript denote partial differentiation.

Thus the displacement field (13) of the first Timpe solution is of the same form as that of the first Boussinesq solution given in (1). The stress function Ψ of the second Timpe solution satisfies the partial differential equation which characterizes the stream function in axisymmetric potential flows of an ideal, incompressible fluid. Upon transformation of (14) into curvilinear coordinates, we obtain

$$\left.\begin{aligned} u_\alpha &= \frac{h_1}{2G}\left[\Psi_\alpha - 4(1-\nu)\frac{1}{f}f_\alpha\Psi\right] \\ u_\beta &= \frac{h_2}{2G}\left[\Psi_\beta - 4(1-\nu)\frac{1}{f}f_\beta\Psi\right] \end{aligned}\right\} \tag{16}$$

and the corresponding field of stress has the curvilinear representation

$$\left.\begin{aligned} \sigma_\alpha &= h_1^2\Psi_{\alpha\alpha} - \frac{h_1}{f}\left[2(2-\nu)h_1 f_\alpha - f\frac{\partial h_1}{\partial\alpha}\right]\Psi_\alpha \\ &\quad - \frac{h_2^2}{h_1 f}\left[2\nu h_1 f_\beta + f\frac{\partial h_1}{\partial\beta}\right]\Psi_\beta \\ &\quad - \frac{4(1-\nu)}{f}\left\{h_1^2\left[f_{\alpha\alpha} - \frac{1}{f}f_\alpha^2\right] \right. \\ &\quad \left. + h_1 f_\alpha\frac{\partial h_1}{\partial\alpha} - \frac{h_2^2}{h_1}f_\beta\frac{\partial h_1}{\partial\beta}\right\}\Psi \\ \sigma_\beta &= h_2^2\Psi_{\beta\beta} - \frac{h_2}{f}\left[2(2-\nu)h_2 f_\beta - f\frac{\partial h_2}{\partial\beta}\right]\Psi_\beta \\ &\quad - \frac{h_1^2}{h_2 f}\left[2\nu h_2 f_\alpha + f\frac{\partial h_2}{\partial\alpha}\right]\Psi_\alpha \\ &\quad - \frac{4(1-\nu)}{f}\left\{h_2^2\left[f_{\beta\beta} - \frac{1}{f}f_\beta^2\right] \right. \\ &\quad \left. + h_2 f_\beta\frac{\partial h_2}{\partial\beta} - \frac{h_1^2}{h_2}f_\alpha\frac{\partial h_2}{\partial\alpha}\right\}\Psi \\ \sigma_\gamma &= \frac{1-2\nu}{f}\left[h_1^2 f_\alpha\Psi_\alpha + h_2^2 f_\beta\Psi_\beta\right] \\ &\quad - \frac{4(1-\nu)}{f^2}\left[h_1^2 f_\alpha^2 + h_2^2 f_\beta^2\right]\Psi \\ \tau_{\alpha\beta} &= h_1 h_2\Psi_{\alpha\beta} - h_2\left[2(1-\nu)\frac{h_1}{f}f_\beta - \frac{\partial h_1}{\partial\beta}\right]\Psi_\alpha \\ &\quad - h_1\left[2(1-\nu)\frac{h_2}{f}f_\alpha - \frac{\partial h_2}{\partial\alpha}\right]\Psi_\beta \\ &\quad - \frac{4(1-\nu)}{f}\left\{h_1 h_2\left[f_{\alpha\beta} - \frac{1}{f}f_\alpha f_\beta\right] \right. \\ &\quad \left. + h_1 f_\beta\frac{\partial h_2}{\partial\alpha} + h_2 f_\alpha\frac{\partial h_1}{\partial\beta}\right\}\Psi \end{aligned}\right\} \tag{17}$$

whereas the second of Eqs. (15) becomes

$$\frac{\partial}{\partial\alpha}\left(\frac{h_1}{fh_2}\Psi_\alpha\right) + \frac{\partial}{\partial\beta}\left(\frac{h_2}{fh_1}\Psi_\beta\right) = 0. \tag{18}$$

A CONNECTION BETWEEN THE SOLUTIONS OF BOUSSINESQ AND TIMPE

We now determine Boussinesq stress functions φ and ψ which are equivalent to given Timpe functions Φ and Ψ in the sense that they generate the same displacement and stress fields. To this end we equate the displacement fields of the complete Boussinesq and Timpe solutions, and obtain with the aid of (1), (2) and (13), (14),

$$\begin{aligned} &\operatorname{grad}(\Phi+\Psi) - \left[\frac{4(1-\nu)}{\rho}\Psi, 0, 0\right] \\ &\qquad = \operatorname{grad}(\varphi+\psi) - [0, 0, 4(1-\nu)\psi]. \end{aligned} \tag{19}$$

Application of the divergence and curl operators to (19) yields

$$\rho\psi_z = \Psi_\rho, \quad \rho\psi_\rho = -\Psi_z. \tag{20}$$

Hence $\psi(\rho, z)$ admits the line integral representation

$$\psi(\rho, z) = \int_{(\rho_0, z_0)}^{(\rho, z)}\left[-\frac{1}{\rho}\Psi_z d\rho + \frac{1}{\rho}\Psi_\rho dz\right], \tag{21}$$

where (ρ_0, z_0) is an arbitrary fixed point of the meridional half-plane $\rho \geq 0$. In view of the second of Eqs. (15), the integral (21) is independent of the path, and hence ψ is single-valued in any simply connected domain throughout which Ψ is regular. Moreover,

$$\nabla^2\psi(\rho, z) = 0. \tag{22}$$

The function ψ corresponds to the velocity potential in the axisymmetric potential flows referred to earlier.

Equating ρ-components and z-components in (19), and making use of (20), we reach

$$\left.\begin{aligned} (\partial/\partial\rho)(\varphi-\Phi) &= \rho\psi_z - z\psi_\rho - [4(1-\nu)/\rho]\Psi, \\ (\partial/\partial z)(\varphi-\Phi) &= (3-4\nu)\psi - z\psi_z - \rho\psi_\rho. \end{aligned}\right\} \tag{23}$$

Equations (23) at once imply the following determination of φ:

$$\begin{aligned} \varphi(\rho, z) &= \Phi(\rho, z) \\ &\quad + \int_{(\rho_0, z_0)}^{(\rho, z)}\left\{\left[\rho\psi_z - z\psi_\rho - \frac{4(1-\nu)}{\rho}\nu\right]d\rho \right. \\ &\quad \left. + [(3-4\nu)\psi - z\psi_z - \rho\psi_\rho]dz\right\}. \end{aligned} \tag{24}$$

We observe, on the basis of (20), that $\varphi(\rho, z)$ is single-valued in any simply connected domain in which $\Psi(\rho, z)$ is regular. Furthermore, direct computation

yields

$$\nabla^2\varphi(\rho, z)=0. \tag{25}$$

Equations (21) and (24) constitute an explicit connection between the two stress-function approaches under consideration.

ACKNOWLEDGMENT

The authors are greatly indebted to a reviewer for several helpful criticisms.

BIBLIOGRAPHY

[1] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity* (Dover Publications, New York, 1944), 4th edition.
[2] E. Butty, *Tratado de Elasticidad* (Centro Estudiantes de Ingenieria, Buenos Aires, 1946).
[3] J. Boussinesq, *Applications des Potentiels* (Gauthier-Villars, Paris, France, 1885).
[4] A. and L. Föppl, *Drang und Zwang* (R. Oldenburg, München und Berlin, 1944), Vol. 2, fourth edition.
[5] B. Galerkin, "Contribution à la solution générale du problème de la théorie de l'élasticité dans le cas de trois dimensions," Compt. rend. **190** (1930).
[6] P. F. Papkovich, "Solution générale des équations différentielles fondamentals d'élasticité exprimée par trois fonctions harmonique," Compt. rend. **195** (1932).
[7] H. Neuber, "Ein neuer Ansatz zur Lösung räumlicher Probleme der Elastizitätstheorie," Z. angew. Math. Mech. **14** (1934).
[8] H. Neuber, *Kerbspannungslehre* (Edwards Brothers, Inc., Ann Arbor, Michigan, 1944).
[9] R. D. Mindlin, "Note on the Galerkin and Papkovich stress functions," Bull. Am. Math. Soc. **42** (1936).
[10] A. Timpe, "Achsensymmetrische Deformation von Umdrehungskörpern," Z. angew. Math. Mech. **4** (1924).
[11] C. Weber, "Achsensymmetrische Deformation von Umdrehungskörpern," Z. angew. Math. Mech. **5** (1925).
[12] M. A. Sadowsky and E. Sternberg, "Stress concentration around a triaxial ellipsoidal cavity," J. Appl. Mech. **16**, No. 2 (1949).
[13] R. D. Mindlin, "Force at a point in the interior of a semi-infinite solid," Physics **7** (1936).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 9 SEPTEMBER, 1951

The Offset Wave-Guide Junction as a Reactive Element*

L. D. SMULLIN AND W. G. GLASS
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received February 23, 1951)

A new type of wave-guide circuit element has been described that has a number of desirable electrical and mechanical properties. These include ease and accuracy of adjustment, and simple analytical expressions for some of the more important properties.

The capacitive junction allows the practical construction of capacitively coupled resonant cavities. Such cavities have a relatively constant band width over their tuning range, as compared with the more commonly used inductively coupled cavities.

INTRODUCTION

THE microwave circuit designer has at his disposal a number of different mechanical forms with which to produce lumbed capacitive or inductive reactances. One can produce a capacitive shunt reactance across a wave guide by means of a tuning screw whose axis is parallel to the electric field lines in the guide, or by means of an iris or partition across the guide (Fig. 1a). Similarly, inductive shunt reactances can be realized by posts across the guide, or by irises (Fig. 1b). Except for the capacitive tuning screw, which can produce only small susceptances, none of these structures is easily variable. Furthermore, the capacitive iris, which requires very narrow gaps in order to produce appreciable susceptances, is difficult to fabricate with any precision.

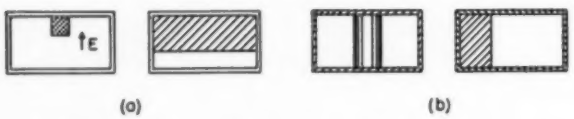


FIG. 1. Conventional forms of capacitive and inductive elements used in wave guide.

The offset wave-guide junctions described in this paper are characterized by a rugged mechanical construction and ease of adjustment. Figure 2 illustrates the three basic forms of these junctions: capacitive, inductive, and resonant. They are made from sections of wave guides having plain flanges, clamped or soldered together at the desired offset. This simple construction allows gaps to be adjusted to extremely close tolerances by means of "feeler gauges," so that, for example, it is relatively easy to make a 0.005-inch capacitive gap in a $1\times\frac{1}{2}\times 0.050$-inch wave guide. Similarly, very narrow resonant gaps can be made without the difficulties from warping that are encountered in conventional irises.

In addition to their mechanical features, the offset junctions have some interesting electrical properties. The capacitive junction turns out to be the equivalent of an infinitely thin asymmetrical capacitive iris and can be described by the same equation. It behaves as a simple shunt capacitance across the wave guide, located in the plane of the junction. The inductive junction, because of asymmetry of the H field on opposite sides of the junction, must be described by a T of inductances.

* This work was supported in part by the Signal Corps, Air Materiel Command, and ONR.

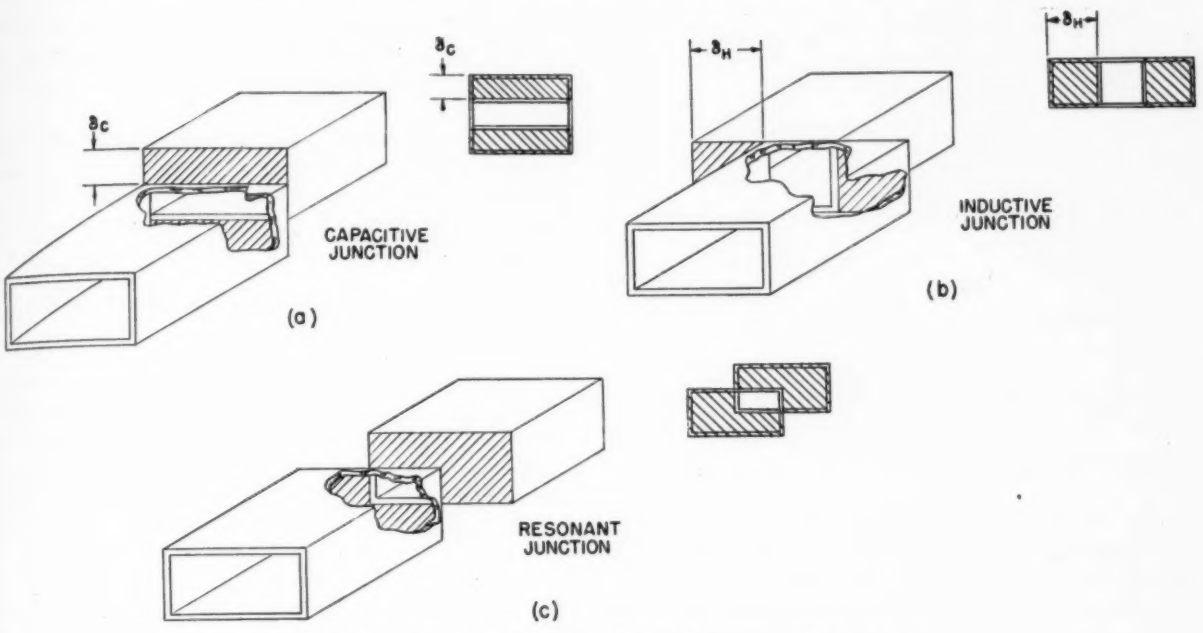


FIG. 2. The three principal forms of the offset wave-guide junction.

The shunt element varies as predicted by the equation for an infinitely thin, asymmetric, inductive iris. Finally, the resonant junction can be described in terms of loci of constant resonant frequency with varying Q. These loci turn out to be hyperbolas, but no expression for the loaded Q has yet been obtained.

CHARACTERISTICS OF OFFSET JUNCTIONS

All measurements were made at wavelengths around 3.3 cm with an RG-52U wave guide. Depending upon the range of susceptance to be determined, measurements were made with either a precision magic T or with a slotted line.

Figure 3 is an admittance plot for the capacitive and inductive junctions, in parallel with a matched termination. The offset distances, δ_H for the inductive and δ_c for the capacitive junction, are indicated along the curves. It can be seen that the points for the capacitive junction lie very closely on the $1+jb$ circle, corresponding to a pure shunt capacitance. This is not true of the points for the inductive junction. A lossless two-terminal-pair network can be described, at a single frequency, by either a T or a π of reactive elements. According to Dicke[1] a wave-guide obstacle may be described as a simple shunt element if it is possible

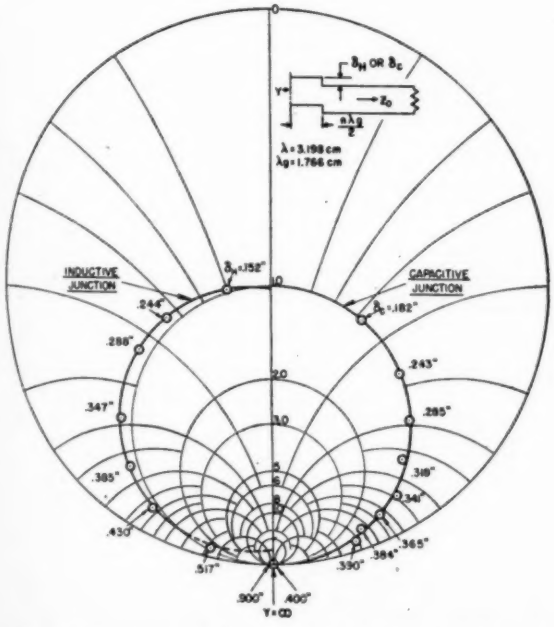


FIG. 3. Admittance *vs* offset (δ_H or δ_c) of the inductive and capacitive junctions, in parallel with Z_0.

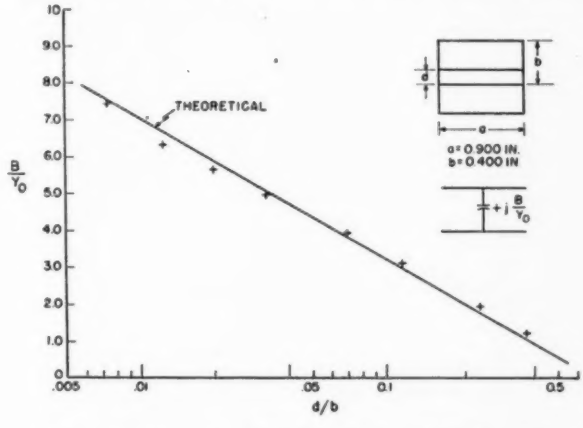


FIG. 4. Susceptance *vs* offset of the capacitive junction, $\lambda = 3.315$ cm.

[1] Montgomery, Dicke, and Purcell, *Principles of Microwave Circuits* (McGraw-Hill Book Company, Inc., New York, 1948), Radiation Laboratory Series, M.I.T., Vol. 8.

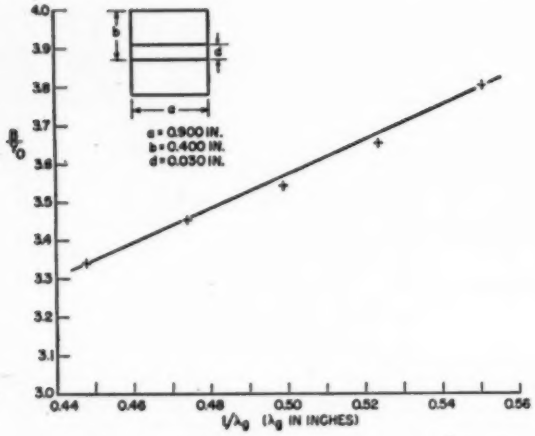


FIG. 5. Susceptance *vs* $1/\lambda_g$ of a capacitive junction.

to make the field in the plane of the obstacle zero by sending two waves of appropriate magnitude and phase from opposite sides. That is, the series elements of the equivalent T circuit vanish. Since the y variation of the electric fields is the same on both sides of the junction and the x variations are the same except for sign, this test is met by the capacitive junction. In the inductive junction, this condition cannot be satisfied because of the reversed variation of H_z on opposite sides of the junction, and the complete T circuit must be used.

Figure 4 is a plot of the normalized shunt susceptance of a capacitive junction, compared with the expression for an asymmetrical iris[1]

$$B/Y_0=(8b/\lambda_g)\ln[\csc(\pi d/2b)],$$

where d is the iris opening, and b the guide height. The agreement is well within the experimental error. Figure 5 is a plot of B/Y_0 *vs* $1/\lambda_g$ for one value of offset; the points represent a straight line, in agreement with the equation.

Figure 6 shows the values of the shunt and series elements of the equivalent T circuit of an inductive junction. The theoretical value of the shunt susceptance of an infinitely thin asymmetrical iris is also plotted[1] in Fig. 6a:

$$B/Y_0=-[(\lambda_g/a)\cot^2(\pi d/2a)][1+\csc^2(\pi d/2a)],$$

where a is the guide width.

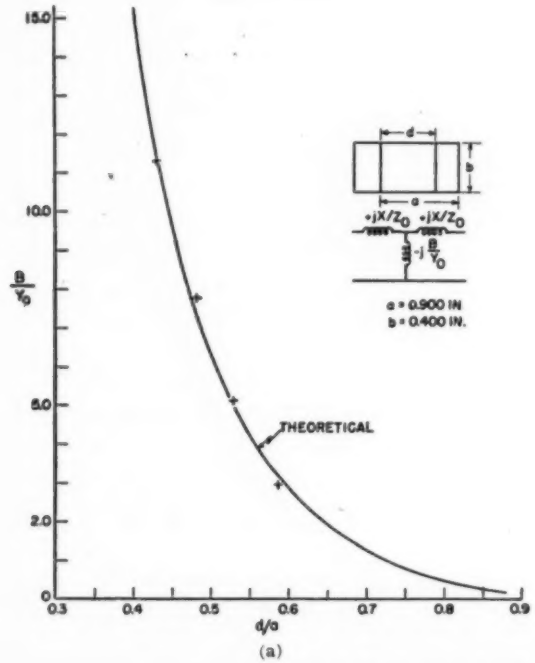
Figure 7 is a plot of the loci of constant resonant wavelength of the resonant junction, in which the guides were displaced diagonally, with their transverse axes parallel. The intercepts at which the vertical opening just vanishes, are very nearly equal to half of the air wavelength, as indicated in the table:

$2a_0'$	λ
3.124 cm	3.136 cm
3.328	3.359
3.532	3.613.

These curves are very closely described by hyperbolas of the form

$$\frac{a'[1-(\lambda/2a')^2]^{\frac{1}{2}}}{a[1-(\lambda/2a)^2]^{\frac{1}{2}}}=\frac{b'}{b}.$$

No extensive measurements of loaded Q of these junctions were made. It is known, however, that Q_{L_2}, the Q of the junction loaded by $2Y_0$, is low; and for a horizontal offset $\delta_H=0.21$, $Q_{L_2}\approx 3$.



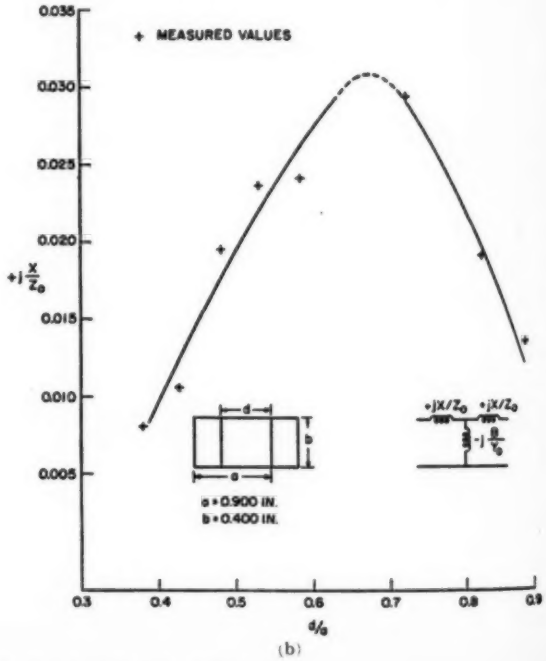


FIG. 6. Equivalent circuit of the inductive junction, $\lambda=3.321$ cm.

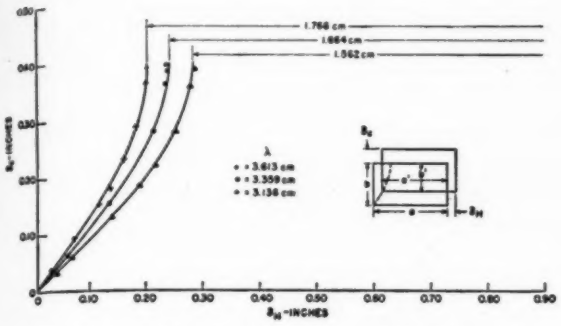


FIG. 7. Loci of constant resonant wavelength for the resonant junction.

APPLICATIONS TO FILTERS

It is to be expected that these junctions will find their greatest usefulness in the design and construction of microwave filters. A typical set of building blocks for a quarter-wave coupled filter is shown in Fig. 8. These are simply wave guides cut to the desired length (approximately $\lambda_g/2$ and $\lambda_g/4$, respectively) and equipped with plane flanges. The theory of multiple cavity filters is well developed,[2] so that, given the characteristics of the elements, one can predict the filter performance with a high degree of precision.

The accuracy with which the elementary resonant cavity can be designed is illustrated in Fig. 9. These curves are for inductively and capacitively coupled tunable cavities. The tuning was accomplished by means of screws at the centers of the respective cavities. The loaded Q was calculated[3] from the equation

$$Q_{L_2}=\frac{(b_0^2+1)\theta_g(\lambda_g/\lambda)^2+b_0}{(1/b_1^2)+(1/b_2^2)+g_c},$$

where b_1 and b_2 are the normalized coupling susceptances, g_c is the normalized conductance of the cavity at its midpoint, and the tuning screw susceptance is

$$2b_0=[(b_1+\tan\theta)/b_1\tan\theta]+(b_2+\tan\theta)/b_2\tan\theta,$$

where

$$\theta=2\pi l/\lambda_g \text{ and } 2l=\text{cavity length}.$$

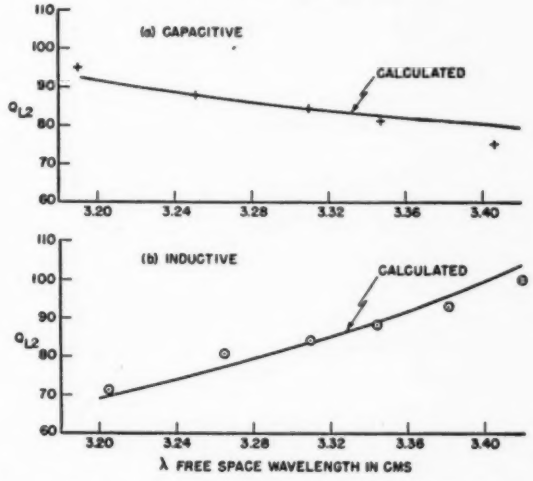


FIG. 9. Q_{L_2} *vs* tuning of capacitively and inductively coupled cavities. (a) $\delta_c=0.005''$, $2l=0.952''$. (b) $\delta_H=0.430''$, $2l=0.798''$, $g_c\ll 1/b_1^2$.

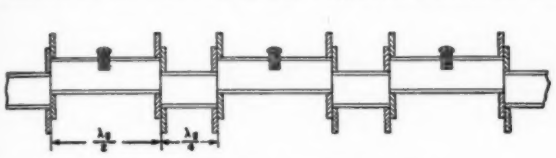


FIG. 8. Multiple cavity filter assembly.

These curves illustrate the advantage of using capacitively coupled cavities for tunable cavities, or filters; the variation in loaded Q with tuning is considerably less than for an inductively coupled cavity.[4]

[2] G. L. Ragan, *Microwave Transmission Circuits* (McGraw-Hill Book Company, Inc., New York, 1948), Radiation Laboratory Series, M.I.T., Vol. 9, Chapters 9 and 10, pp. 540–715.

[3] L. D. Smullin and C. G. Montgomery, *Microwave Duplexers* (McGraw-Hill Book Company, Inc., New York, 1948), Radiation Laboratory Series, M.I.T., Vol. 14.

[4] L. D. Smullin, Technical Report No. 106, Research Laboratory of Electronics, M.I.T. (1949); Abstract, Proc. Inst. Radio Engrs. **37**, 12 (1949).

Mathematical Data for Electron Drain on Positive Ion Sources

S. E. RAUCH
Department of Mathematics, University of California, Santa Barbara College
(Received February 28, 1951)

A study of the motion of electrons moving in the presence of crossed electric and magnetic fields is presented. Special electromagnetic regions are selected for discussion which assist in the controlling of electron discharge on positive ion sources when used with large mass-spectrograph units. The article compares the components of motion parallel and perpendicular to the magnetic field in the following regions: uniform crossed electric and magnetic fields; electric field produced by nonparallel plane condenser plates, uniform H parallel to one plate; electric field produced by coaxial circular cylindrical condensers, uniform magnetic field normal to axis of cylinders; uniform electric field, circularly bowed magnetic field. The components of motion are expressed in terms of mathematical functions, and where it is deemed useful, additional results are presented in graphical form so that the conclusions can be applied with a minimum of calculations.

INTRODUCTION

IN the operation of positive ion sources for nuclear investigations, such as in the electromagnetic separation of isotopes, an important engineering problem arises in controlling the loss of electrical power due to electrons. The positive ion source is maintained at a high positive voltage but is surrounded by a ground shield to prevent electrical field spread into the high vacuum magnetic region. Electrons multiply by familiar processes in the gaseous region between the positive ion chamber and its ground shield. The electrons have a component of motion both parallel and perpendicular to the magnetic field, but unfortunately the component perpendicular to the magnetic field in the direction of the equipotentials can be very large. Effective engineering designs attempt to maintain minimum electron paths and velocities because in the large units the drain of electrons upon the positive source absorbs large amounts of electrical energy, producing excessive heating, and in serious instances destroys metallic surfaces. Various types of electron traps are employed in practice to assist the reduction of electron multiplication in such regions.

The available literature concerning the paths of charged particles in crossed electric and magnetic fields is concerned with application where the initial energies of the charged particles tre important in determining the type of path. Consequently, cycloidal and helical paths have been studied fully.[1–5] The previous studies do not offer directly results that can be used for electron trap regions. In the present discussion the initial energies of electrons are negligible in comparison with the energy of the electric field or the energy gained by the electron in the field. Thus the electrons have a drift component in the direction defined by $\mathbf{E}\times\mathbf{H}$, that is, along equipotentials, with only minor fluctuations due to cycloidal variation. In addition the positive ion space charge is considered negligible for the regions being studied.

In the following work, it is assumed that E is in esu, V is in practical volts, H is in gauss, and distances are in centimeters.

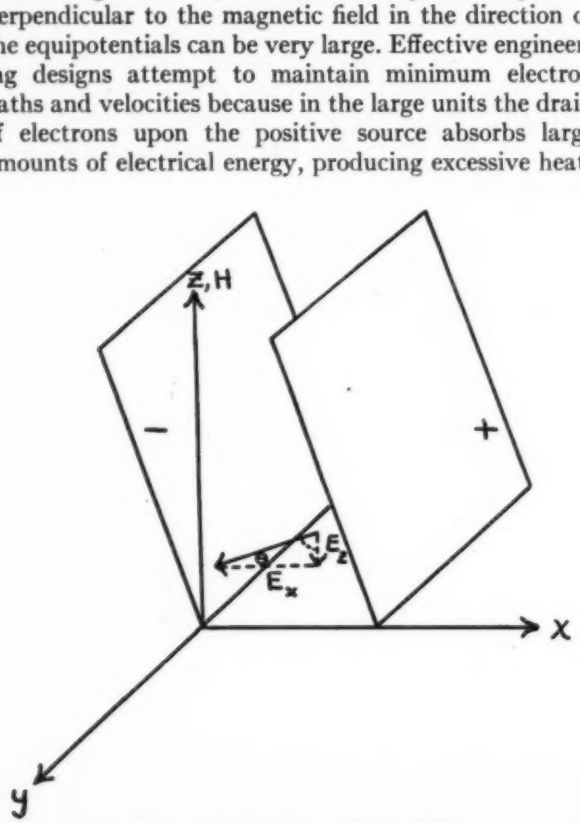


FIG. 1. Uniform crossed E and H.

I. UNIFORM CROSSED E AND H

Let the uniform magnetic field H be parallel to the z axis, the parallel condenser plates be perpendicular to the $x-z$ plane, the uniform electric field E be inclined at an angle θ with the x axis, as shown in Fig. 1.

The equations of motion are as follows:

$$m\ddot{x}=eE_x-e\dot{y}H/C,$$
$$m\ddot{y}=e\dot{x}H/C,$$
$$m\ddot{z}=eE_z.$$

The solution of the system of differential equations is quite simple upon assuming initial coordinates and velocities to be zero. Then the parametric form of the

[1] N. D. Coggeshall and M. Muskat, Phys. Rev. **66**, 187 (1944).
[2] N. D. Coggeshall and M. Muskat, Phys. Rev. **70**, 270 (1946).
[3] A. J. Dempster, Proc. Am. Phil. Soc. **75**, 755 (1935).
[4] F. W. Aston, Nature **137**, 357 (1936).
[5] R. G. E. Hutter, Phys. Rev. **67**, 248 (1945).

solution becomes

$$y=(CE_z/H)t,$$
$$z=(eE_z/2m)t^2.$$

Thus the electrons travel along equipotentials with energies corresponding to the average drift velocity $U=C\cdot E_z/H$.

Let the separation between the parallel plane condenser plates be s. The distance y that an electron travels in the drift direction in traversing a distance z parallel to H is

$$y=58.2\ (\cos\theta/H)(E_z/\sin\theta)^{\frac{1}{2}}.$$

Example: $E=2$ esu, $H=3000$ gauss, $\theta=5°$, then $y=0.024(z)^{\frac{1}{2}}$ cm.

The drift distance y for an electron to travel from the negative condenser plate to the positively charged plates is given by

$$y=(C/H)(mV/150e)^{\frac{1}{2}}\cot\theta$$
$$=[(V)^{\frac{1}{2}}/H]\cdot 3.37\cot\theta.$$

Example: $V=10{,}000$ volts, $H=3000$ gauss, then $y=0.112\cot\theta$ cm.

II. E PRODUCED BY NONPARALLEL PLANE CONDENSER, UNIFORM H PARALLEL TO ONE CONDENSER PLATE

Let H be parallel to the Z axis, the condenser plates be perpendicular to the $x-Z$ plane and intersecting at an angle β, the negatively charged plate be parallel to Z axis. At the time $t=0$, let the electron be located at (r, θ_0), where r and θ are polar coordinates in the $x-z$ plane, as shown in Fig. 2.

For the above configuration,

$$E=V/300r\beta,\ E_x=V\cdot\sin\theta/300r\beta,\ E_z=-V\cdot\cos\theta/300r\beta.$$

The distance y that the electron travels parallel to the equipotentials in order to reach the positive surface is given by

$$y=\frac{C}{2H}\left(\frac{mV}{150e\beta}\right)^{\frac{1}{2}}\int_{\pi/2-\beta}^{\theta_0}\frac{\tan\theta d\theta}{(\theta_0-\theta)^{\frac{1}{2}}},$$

$$y=\frac{C}{2H}\left(\frac{mV}{150e\beta}\right)^{\frac{1}{2}}\left\{\int_{\pi/2-\beta}^{\theta_0}\frac{\tan\theta-\tan\theta_0}{(\theta_0-\theta)^{\frac{1}{2}}}d\theta+2\tan\theta_0(\theta_0+\beta-\pi/2)^{\frac{1}{2}}\right\}.$$

It is easily shown that for $\theta<\theta_0$

$$\left|\int_\theta^{\theta_0}\left(\frac{\tan\theta-\tan\theta_0}{(\theta_0-\theta)^{\frac{1}{2}}}\right)d\theta\right|<\tfrac{1}{2}(\tan\theta_0-\tan\theta)(\theta_0-\theta)^{\frac{1}{2}};$$

therefore,

$$y<\frac{C}{H}\left(\frac{mV}{150e}\right)^{\frac{1}{2}}\left(\frac{\theta_0+\beta-\pi/2}{\beta}\right)^{\frac{1}{2}}\tan\theta_0,$$

$$y>\frac{C}{2H}\left(\frac{mV}{150e}\right)^{\frac{1}{2}}\left(\frac{\cot\beta+3\tan\theta_0}{2}\right)\left(\frac{\theta_0+\beta-\pi/2}{\beta}\right)^{\frac{1}{2}}.$$

For small β,

$$y\sim\frac{C}{H}\left(\frac{mV}{150e}\right)^{\frac{1}{2}}\left(\frac{\theta_0+\beta-\pi/2}{\beta}\right)^{\frac{1}{2}}\tan\theta_0.$$

Define $\phi=\frac{1}{2}\pi-\theta$, $\phi_0=\frac{1}{2}\pi-\theta_0$, then

$$y\sim\frac{C}{H}\left(\frac{mV}{150e}\right)^{\frac{1}{2}}\cot\phi_0\left(1-\frac{\phi_0}{\beta}\right)^{\frac{1}{2}},$$

$$y\sim\frac{(V)^{\frac{1}{2}}}{H}0.275\cot\phi_0\left(1-\frac{\phi_0}{2\beta}\right).$$

Example: let $H=3000$ gauss, $V=10{,}000$ volts, $\phi_0/2\beta=0.5$, then $y\sim0.007\cdot\cot\phi_0$ cm.

III. E PRODUCED BY COAXIAL CIRCULAR CONDENSER, UNIFORM H NORMAL TO AXIS OF CYLINDERS

Let the axis of the cylinders be coincident with the y axis, the radii of circular cylinders be a and b, $a<b$, (r, θ_0) be the location of the electron at $t=0$, where r and θ are polar coordinates in the $x-z$ plane as shown in Fig. 3. The electric field is described by $E=V/300r\cdot\ln(b/a)$. For the following let $K=V/300\cdot\ln(b/a)$. The type of path followed by the electron is dependent upon the positive charge being located on the outer or inner cylinder of the condenser. The two possible cases are described separately in the following.

A. Positive Charge on Outer Cylindrical Surface of Condenser

The distance y traveled in the drift direction for the electron to move from θ_0 to θ along a line of the magnetic

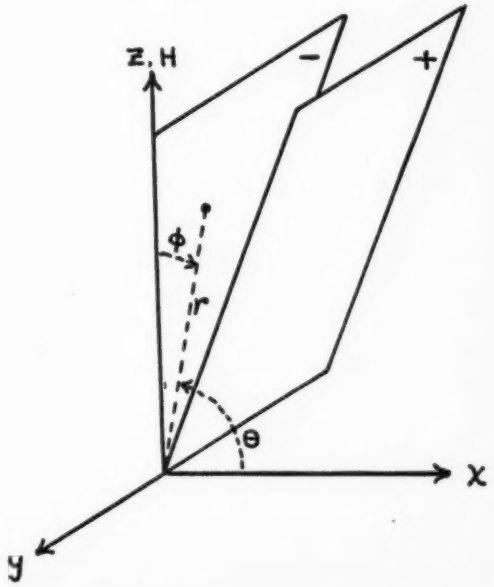


FIG. 2. E produced by nonparallel plane condenser, uniform H parallel to one condenser plate.

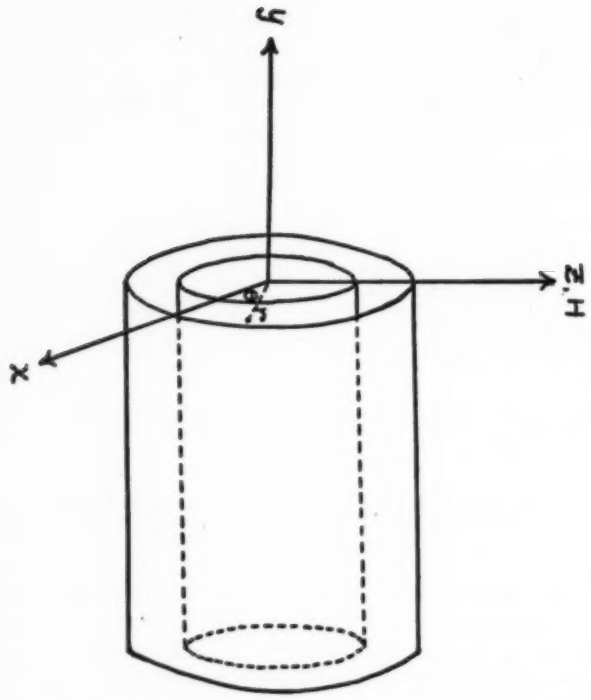


FIG. 3. E produced by coaxial cylindrical condenser, uniform H normal to axis of cylinders.

field is easily shown to be

$$y=\frac{C}{H}\left(\frac{mK}{2e}\right)^{\frac{1}{2}}\int_{\theta_0}^{\theta}\frac{d\theta}{\left(\ln\dfrac{\sec\theta}{\sec\theta_0}\right)^{\frac{1}{2}}},$$

or for the purposes of computation it becomes

$$y=\frac{(V)^{\frac{1}{2}}}{H}\cdot\frac{1}{[\ln(b/a)]^{\frac{1}{2}}}\cdot I_3,$$

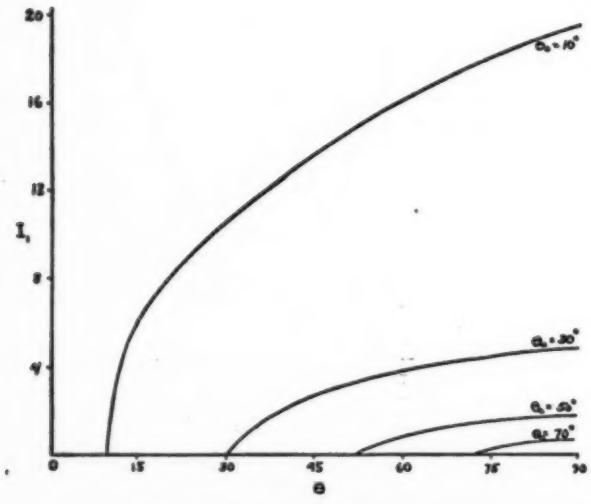


FIG. 4. I_1 as a function of θ_0 and θ.

where

$$I_3=1.686\left\{\int_{\theta_0}^{\theta}\left[\frac{1}{\left(\ln\dfrac{\sec\theta}{\sec\theta_0}\right)^{\frac{1}{2}}}-\frac{\cot\theta_0}{(\theta-\theta_0)^{\frac{1}{2}}}\right]d\theta+2\cot\theta_0(\theta-\theta_0)^{\frac{1}{2}}\right\}.$$

I_1 as a function of θ_0 and θ is presented in Fig. 4. Example: let $H=3000$ gauss, $V=10{,}000$ volts, $b/a=1.2$, then $y\sim 0.08\cdot I_1$; if $\theta_0=30°$, $\theta=45°$, $I_1=2.8$ and $y=0.2$ cm.

B. Positive Charge on Inner Cylindrical Surface of Condenser

1. $a<x<b$.

In this region the electron cannot reach a positive surface because of its restriction to move parallel to the magnetic field for its component of motion parallel to the z axis. Thus the electron oscillates between $+\theta_0$ and $-\theta_0$ as it drifts in the y direction.

The period of oscillation is given by

$$T=4x\left(\frac{m}{eK}\right)^{\frac{1}{2}}\int_0^{u_0}\frac{du}{\left(\ln\dfrac{1+u_0^2}{1+u^2}\right)^{\frac{1}{2}}},$$

where $u=\tan\theta$, $u_0=\tan\theta_0$. Upon letting

$$I_T=0.954\int_0^{u_0}\frac{du}{\left(\ln\dfrac{1+u_0^2}{1+u^2}\right)^{\frac{1}{2}}}$$

then,

$$T=x\left(\frac{\ln(b/a)}{V}\right)^{\frac{1}{2}}\cdot I_T\cdot 10^{-7}.$$

The wavelength of oscillation is given by

$$\lambda=\frac{4c}{H}\left(\frac{mK}{e}\right)^{\frac{1}{2}}\int_0^{u_0}\frac{du}{(1+u^2)\left(\ln\dfrac{1+u_0^2}{1+u^2}\right)^{\frac{1}{2}}},$$

and if

$$I_\lambda=0.954\int_0^{u_0}\frac{du}{(1+u^2)\left(\ln\dfrac{1+u_0^2}{1+u^2}\right)^{\frac{1}{2}}},$$

then

$$\lambda=\frac{10}{H}\left(\frac{V}{\ln(b/a)}\right)^{\frac{1}{2}}\cdot I_\lambda.$$

The average drift velocity in the direction of $\mathbf{E}\times\mathbf{H}$ becomes

$$v=\lambda/T,$$

$$v=[V/xH\ln(b/a)]\cdot I_v\cdot 10^8,$$

where

$$I_v=I_\lambda/I_T.$$

Figure 5 presents I_T, I_λ, I_v as a function of θ_0. Example: let $H=3000$ gauss, $V=10{,}000$ volts, $b/a=1.2$, $\theta_0=30°$, $I_T=1.75$, $I_\lambda=1.45$, $I_v=0.86$, and $T=7.5\cdot x\cdot 10^{-10}$ sec, $\lambda=1.2$ cm, $v=1.6/x\cdot 10^9$ cm/sec.

2. $0<x<a$.

For this range of x values the electron can strike a positive surface after traveling the distance y in the drift direction, where

$$y=\frac{C}{H}\left(\frac{mK}{2e}\right)^{\frac{1}{2}}\int_{\theta_0}^{\theta}\frac{d\theta}{\left(\ln\dfrac{\sec\theta_0}{\sec\theta}\right)^{\frac{1}{2}}}.$$

For convenience of computation the above is written in the form,

$$y=(V^{\frac{1}{2}}/H)\cdot I_2,$$

where

$$I_2=\frac{1.686}{[\ln(b/a)]^{\frac{1}{2}}}\left\{\int_{\theta_0}^{\theta}\left[\frac{2\operatorname{ctn}\theta_0}{(\theta_0-\theta)^{\frac{1}{2}}}-\frac{1}{\left(\ln\dfrac{\sec\theta_0}{\sec\theta}\right)^{\frac{1}{2}}}\right]d\theta+4\operatorname{ctn}\theta_0(\theta_0-\theta)^{\frac{1}{2}}\right\}.$$

In Fig. 6, I_2 is presented as a function of b/a and x/a with the additional requirement that the electron is initially at the outer cylinder and traverses the region to the inner positive cylinder.

Example: let $H=3000$ gauss, $V=10{,}000$ volts, $b/a=1.4$, $x/a=0.6$; then $I_2=1.65$, and $y=0.055$ cm.

IV. UNIFORM E, CIRCULARLY BOWED H

Consider an electric field produced by a parallel plane condenser whose plates are perpendicular to the x axis. Let the $x-y$ plane be the plane of symmetry for the circularly bowed magnetic field, the magnetic field not be a function of the y variable, the electron be located at (r, θ_0) at time $t=0$, where r is the constant radius of curvature of the magnetic field and θ is the angle of inclination of the radius vector to the electron above the $x-y$ plane, as is shown in Fig. 7.

The motion of the electron is the resultant of the following: acceleration along H towards the positive surface, $m\cdot r\cdot\ddot{\theta}=cE\cdot\sin\theta$, a cycloidal drift velocity given by $y=(cE/H)\cos\theta$. The distance y traveled in the drift

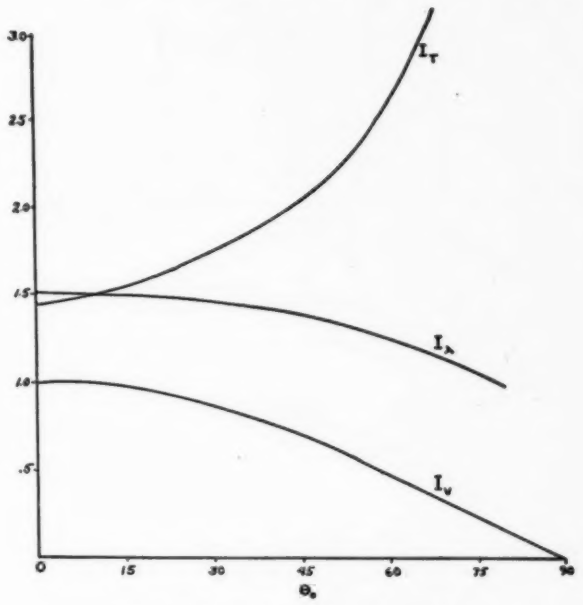


FIG. 5. I_T, I, I_v as a function of θ_0.

direction for the electron to move from θ_0 to θ is given by

$$y=\frac{C}{H}(Erm/e)^{\frac{1}{2}}\int_{\theta_0}^{\theta}\frac{\cos\theta d\theta}{[2(\cos\theta_0-\cos\theta)]^{\frac{1}{2}}}.$$

It is convenient to make the following definitions, $\alpha=180°-\theta$, $\alpha_0=180°-\theta_0$, $\sin\alpha_0/2=k$, $\sin\alpha/2=k\cdot\sin\phi$. Then, the previous statement for y can be expressed as

$$y=\frac{C}{H}(Erm/e)^{\frac{1}{2}}\left\{2k^2\int_{\phi}^{\pi/2}\frac{\sin^2\phi d\phi}{(1-k^2\sin^2\phi)^{\frac{1}{2}}}+F(\phi,k)-F(\pi/2,k)\right\},$$

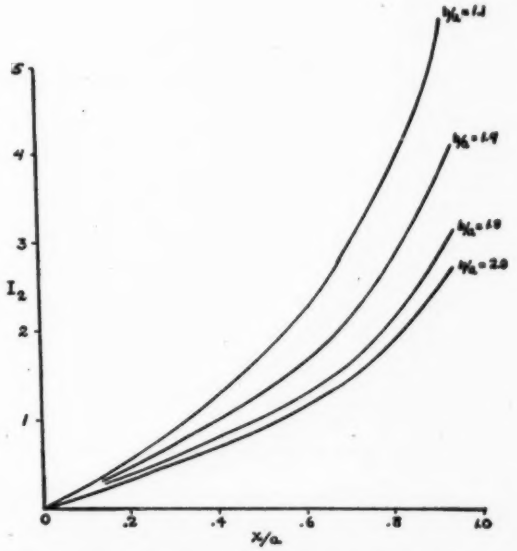


FIG. 6. I_2 as a function of b/a and x/a for an electron to move from outer to inner cylinder.

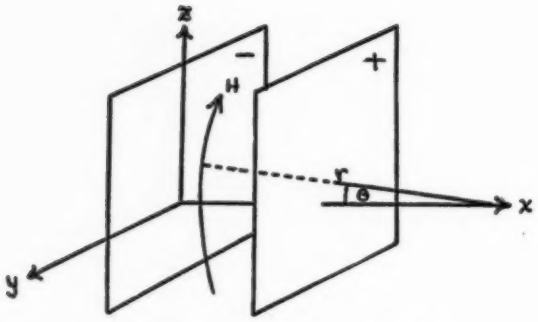


FIG. 7. Uniform E, circularly bowed H.

where

$$F(\phi, k)=\int_0^{\phi=\arcsin[(\cos\theta/2)/k]} \frac{d\phi}{(1-k^2\sin^2\phi)^{\frac{1}{2}}}.$$

Thus

$$|y|<\frac{C}{H}(Erm/e)^{\frac{1}{2}}\{F(\pi/2, k)-F(\phi, k)\}.$$

For the physical applications, r is comparatively large and θ is small. As $\theta\to 0$, $\phi\to\pi/2$, and thus

$$y\sim\frac{C}{H}(Erm/e)^{\frac{1}{2}}\{(\pi/2, k)-F(\phi, k)\},$$

$$y\sim[(E)^{\frac{1}{2}}/H\cdot F_3]\cdot 10^3$$

$$F_3=C(mr/e)^{\frac{1}{2}}\{F(\pi/2, k)-F(\phi, k)\}\cdot 10^{-3}.$$

Figure 8 presents F_3 as a function of r and z_0, for the electron to travel from $z=z_0$ to $z=25$ cm. Example: let $E=10$ esu, $H=3000$ gauss, $r=1200$ cm, $z_0=5$ cm; then $F_3=(3.20)\cdot 10^3$, $y\sim 3.0$ cm.

DISCUSSION

The primary concern in controlling the electron drain upon the positive ion chamber is to minimize the length of electron paths, to minimize the gas pressure between the chamber and ground shields, and to distribute the electron drain over a large surface area in order to prevent localized high temperatures. Most electron traps can be made very effective in making the drift distance in the direction $\mathbf{E}\times\mathbf{H}$ negligible, but large distances parallel to the magnetic field continue to exist. If the distances between electrostatic plates are decreased, it is then found that the gas pressure in the region rises rapidly. Unfortunately the multiplication of electrons is more sensitive to gas pressure than to small changes in distances. Perforations in the ground shields help to remove the gas liberated from the metallic surfaces but the capacity of the vacuum system limits the amount of decrease of distance between plates.

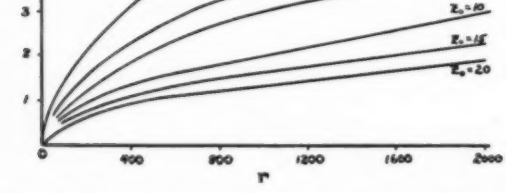


FIG. 8. F_3 as a function of r and z_0 for an electron to travel from $z=z_0$ to $z=25$ cm.

Figure 9 displays the typical electron trap regions. Figure 9a permits long paths parallel to H. This can be remedied only by inclining the plates at a large angle with H. Furthermore, it concentrates the drain on either the top or bottom of the positive ion chamber. Figure 9b has the advantage of spreading the drain over larger areas but the disadvantage of allowing multiplication of electrons in the median plane regions is serious. Figure 9c simplifies the ground shield construction, but it has the same weaknesses as Fig. 9b. Figure 9d provides a region of electron oscillation without permitting electrons to reach a positive surface; however, it can be used with short lengths parallel to $\mathbf{E}\times\mathbf{H}$. The most effective application has been found with the type in Fig. 9d placed in a scatter arrangement on the sides of the positive ion block. The purpose of the scattering is twofold, to spread the electron drain over a large area and to have successive traps remove the electrons that escaped from the preceding trap. The type in Fig. 9d has a simple engineering advantage in that the semicylindrical volumes of metal can be easily placed on the vertical sides of the positive ion chamber.

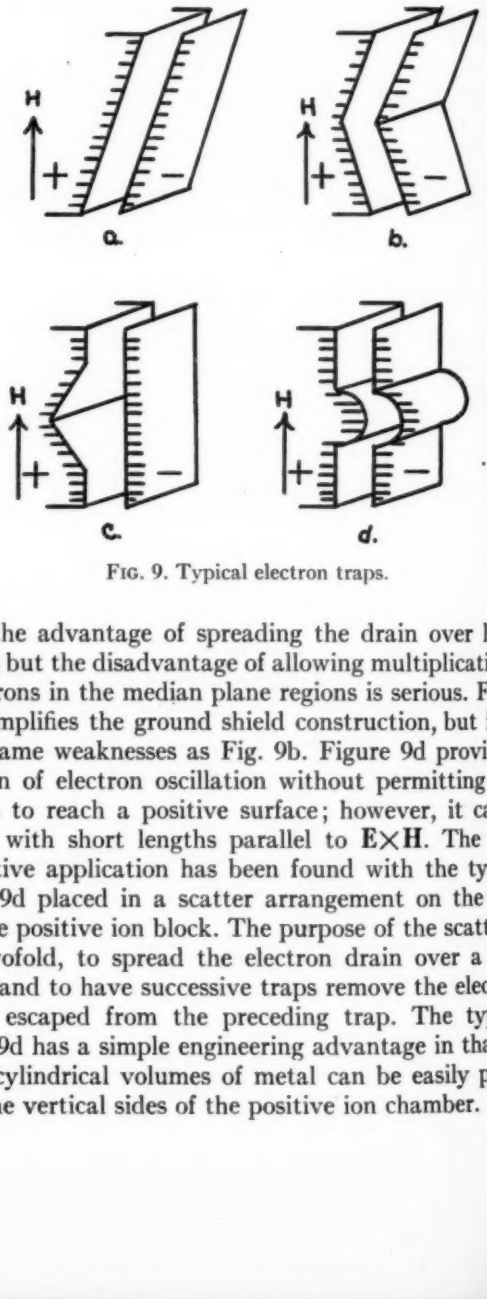


FIG. 9. Typical electron traps.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 9 SEPTEMBER, 1951

Arcing at Electrical Contacts on Closure. Part II. The Initiation of an Arc

L. H. GERMER
Bell Telephone Laboratories, Inc., Murray Hill, New Jersey
(Received March 26, 1951)

The capacity of the plates of an oscilloscope charged to 35 or 40 volts is discharged repeatedly by approaching electrodes of carbon, active silver, and inactive silver. Facts about the discharges, which are arcs of very short duration, are inferred from resulting open circuit potentials and calculated electrode separations.

The separation at the first arc varies in different experiments but corresponds on the average to a nominal electric field of 0.6×10^6 volts/cm for carbon or active silver and to 2×10^6 volts/cm for inactive silver. Each arc is initiated by a very small number of field emission electrons. The hypothesis that a single electron may perhaps be sufficient is consistent with observations at later stages of each closure when the electrodes are closer and the field much higher.

The earlier observation, that the potential across a short arc is constant and independent of current, is not true if the arc time is sufficiently short. For active silver a time comparable with 2×10^{-8} sec is required to establish the steady arc voltage characteristic of later stages of arcs which last longer than this. The initial time during which the potential is decreasing toward its final steady value is 100 times the transit time of a silver ion across the gap.

IF a capacity which is discharged by closing contacts is very small, phenomena can be observed which are not detectable in circuits in which the capacity is large. It is convenient to use for the small capacity that of oscilloscope plates, the experimental circuit consisting then of the contacts in series with the capacity of these plates and the small inductance of the lead wires connecting the plates to the contacts. This circuit is partially isolated from the power supply by a very large inductance, or a large resistance. When the contacts are brought together, the oscilloscope plate capacity is discharged; this discharge undoubtedly takes place in exactly the same way when the inductance or resistance in the principal circuit is not large, but under such conditions the small capacity is recharged so quickly that no precise observation regarding its discharge can be made and even the occurrence of the discharge may escape detection. Tests have been carried out by oscilloscopic observation of the potential across closing contacts which are connected to a relatively large charged capacity through an inductance of large size. The circuit is just that represented in Fig. 16 of an earlier paper,[1] the only difference being the unusual size of the inductance.

The size of this inductance makes it necessary to use a different mechanism for opening and closing the contacts. Because of the largeness of the inductance, 12 henries in some of the tests, the large capacity is not completely discharged in 1/120 sec; and in consequence, the contacts cannot be operated at 60 cycles, as they were in the earlier experiments. They are opened and closed by the same speaker unit, but in place of 60-cycle operation the unit is actuated by a small direct current potential applied and removed at regular intervals by an adjustable multivibrator. The period is 1 or 2 seconds, which is sufficiently long for the capacity to be completely discharged at each closure.

In these tests the contact velocity at closure is required, and an inconvenience attendant upon operation with the multivibrator is the fact that this velocity is not simply determined by observing the amplitude of the driving rod. It has to be found rather laboriously by measuring on the oscilloscope screen the time required for the contacts to close from various separations. The separations are changed by moving the fixed electrode, leaving the driving mechanism and the moving electrode unaltered. The position of the fixed electrode is measured from an arbitrary zero point by a microscope with a filar micrometer eyepiece. For each of a number of

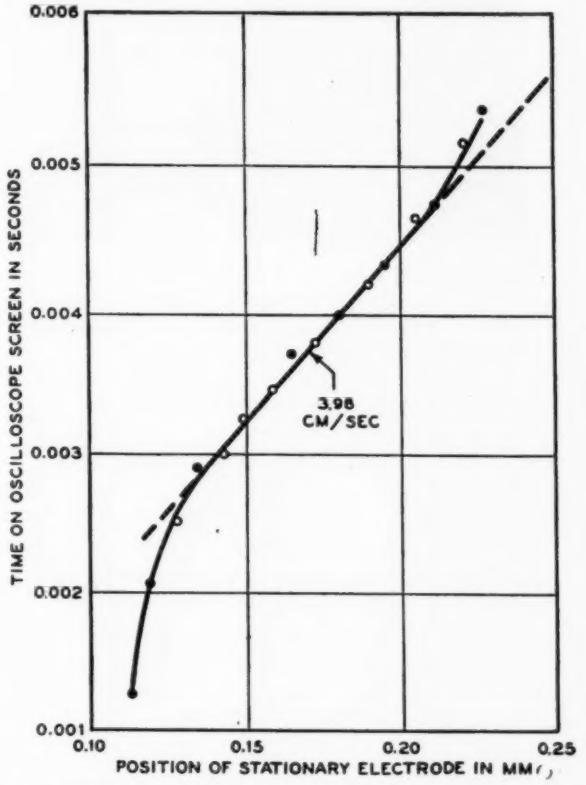


FIG. 1. Plot from which contact velocity at closure is determined. Abscissas represent position of the stationary electrode from an arbitrary zero point, and ordinates time on the oscilloscope screen from the tripping of the sweep until the contacts close.

[1] L. H. Germer and F. E. Haworth, J. Appl. Phys. **20**, 1085 (1949).

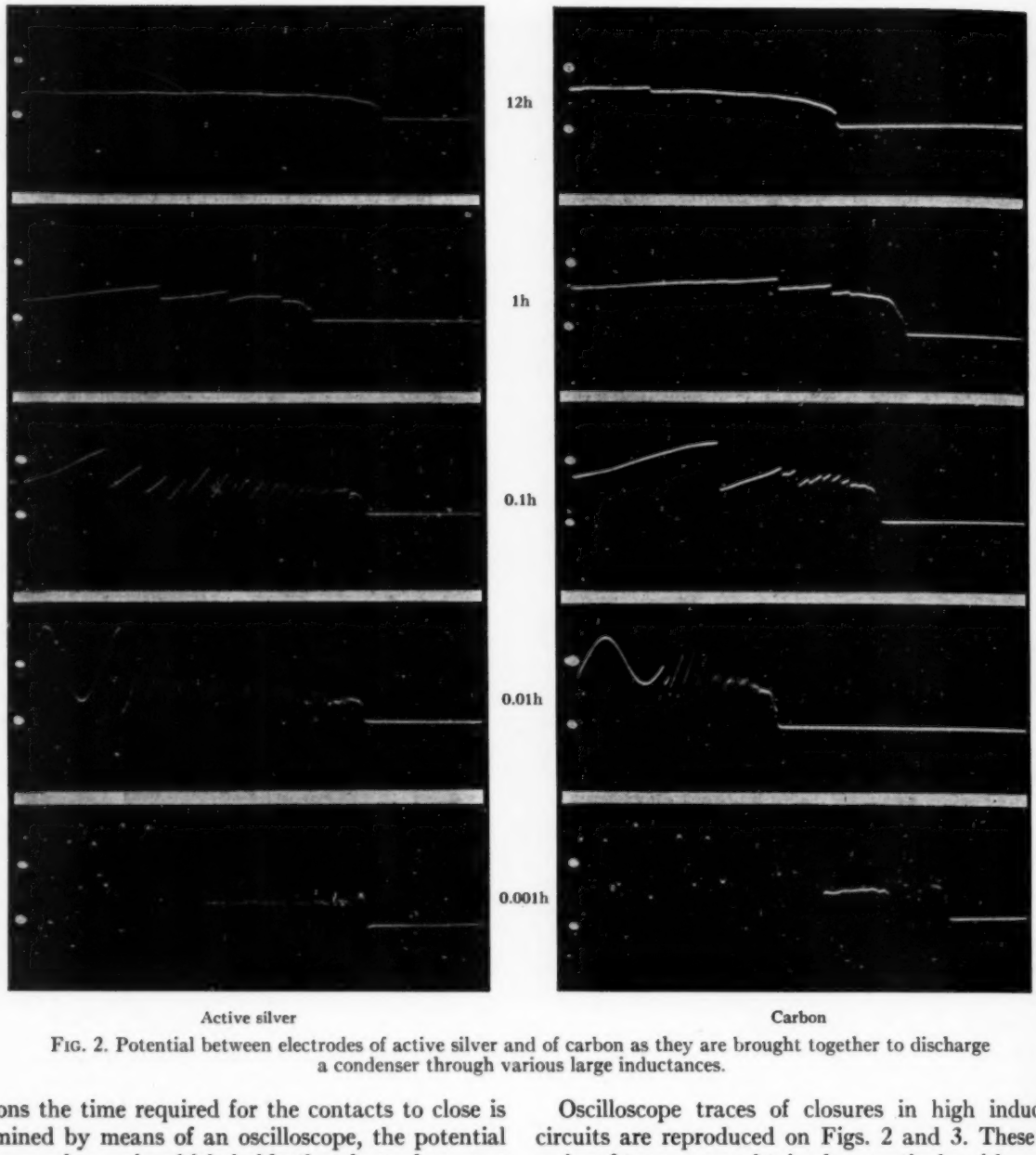


FIG. 2. Potential between electrodes of active silver and of carbon as they are brought together to discharge a condenser through various large inductances.

positions the time required for the contacts to close is determined by means of an oscilloscope, the potential on the speaker unit which holds the electrodes apart being removed simultaneously with the tripping of the sweep. On Fig. 1 are plotted values of time to contact closure, as measured on the oscilloscope screen, against position of the stationary electrode in the field of the microscope from an arbitrary zero. After measurements of this sort have been completed, experiments are carried out with the stationary electrode located at a position represented by an ordinate on this figure which lies on the straight part of the curve. The velocity of the contacts at closure is then known to be the reciprocal of the slope of the curve. For all of the tests reported below the voltage which the multivibrator applies to the speaker unit was adjusted until measurement of the velocity showed that the contacts were closing at 4.0 cm/sec.

Oscilloscope traces of closures in high inductance circuits are reproduced on Figs. 2 and 3. These three series of traces were obtained respectively with contacts of silver activated[2] with *d*-limonene, with contacts of carbon, and with contacts of inactive silver. In each experiment the main capacity which is discharged is $C=10^{-8}$ f, the length of the oscilloscope trace is 22×10^{-4} sec, and the relative velocity of the contacts when they touch is 4.0 cm/sec. The potential difference across the condenser before it is discharged is $V_0=40$ volts for the experiments with carbon electrodes and $V_0=35$ volts for the experiments with silver electrodes. The effective capacity of the oscilloscope plates together with their leads is $C'=40\times10^{-12}$ f; and when this is discharged across the contacts, the charge flows through a total

[2] The first section of this paper is "Part I. Dependence upon surface conditions and circuit parameters," L. H. Germer, J. Appl. Phys. **22**, 955 (1951).

circuit inductance which has been measured to be about $L'=0.2\times10^{-6}$ h.

The traces of Fig. 2 are intended to be representative of several hundred photographed patterns obtained with electrodes of active silver and carbon. With electrodes of inactive silver, however, about half of all photographed traces are like the left trace of Fig. 1A of reference 2 without any feature whatsoever; and the patterns of Fig. 3 have been chosen to be representative of inactive silver after eliminating these. No pattern without features has ever been observed for active silver or for carbon.

The initial sudden drop in potential at the beginning of each trace of Figs. 2 and 3 represents the partial discharge of the oscilloscope plate capacity, and subsequent abrupt small drops which occur before final closure represent additional partial discharges of this capacity. Between these partial discharges one can observe, perhaps most readily in the 1-h and 0.1-h traces of Fig. 2, the recharging of the oscilloscope plates from the much larger capacity which is located beyond the large inductance. For the smaller values of inductance the oscilloscope plate capacity may be charged above its original potential before another discharge occurs, or several cycles of free oscillation may even take place before another discharge. The period of these oscillations is utilized to determine that the oscilloscope plate capacity is 40×10^{-12} f, since this period is determined by this capacity and by the large inductance of known size.

Various features of Figs. 2 and 3 are taken up one at a time in the following paragraphs with as much of their interpretation as is understood at this time. There are so many striking likenesses between the two sets of traces of Fig. 2 that they can be considered together almost all the time, but these traces are markedly different from those of Fig. 3. This observation is in line with previously noted similarities between the arcing behavior of contacts of active silver and of carbon, and with contrasts between active silver and inactive silver.

DEVELOPMENT OF THE LARGE ARC

For the lower values of inductance the traces of Fig. 2 record many successive partial discharges of the oscilloscope plate capacity. The current through the inductance is increasing while these discharges are going on, or rather for that portion of the time for which the voltage across the contacts is below the decreasing voltage on the large condenser. It is to be expected that, if this current finally reaches the minimum arc current,[2] 0.03 amp, a steady arc will be established across the contacts drawing its current from the large condenser through the large inductance. That an arc would probably not be established for an inductance of 0.01h is clear from the fact that if such an arc were to occur, its maximum current $(V_0-v)(C/L)^{\frac{1}{2}}$, where V_0 is the initial voltage across the contacts and v is the arc voltage, would be slightly below the minimum arc current of 0.03 amp; for active silver this works out to be 0.024 amp and for carbon, on the assumption that $v=21$ volts, to be 0.019 amp. For an inductance of 0.001h this expression for maximum current gives 0.076 amp for silver and 0.060 amp for carbon. Just what maximum current could be reached by a large number of successive partial discharges of the oscilloscope plate capacity cannot be worked out because it depends upon the details of these discharges, but this maximum current is probably not very much below the value $(V_0-v)(C/L)^{\frac{1}{2}}$. Looking at the traces of Fig. 2, one observes that there is no arc that is fed directly from the larger capacity for an inductance of 0.01h, but that such arcs are established for both active silver and for carbon when the inductance is 0.001h, the two lowest traces of the figure. The arc for $L=0.001$h lasts a longer time for active silver

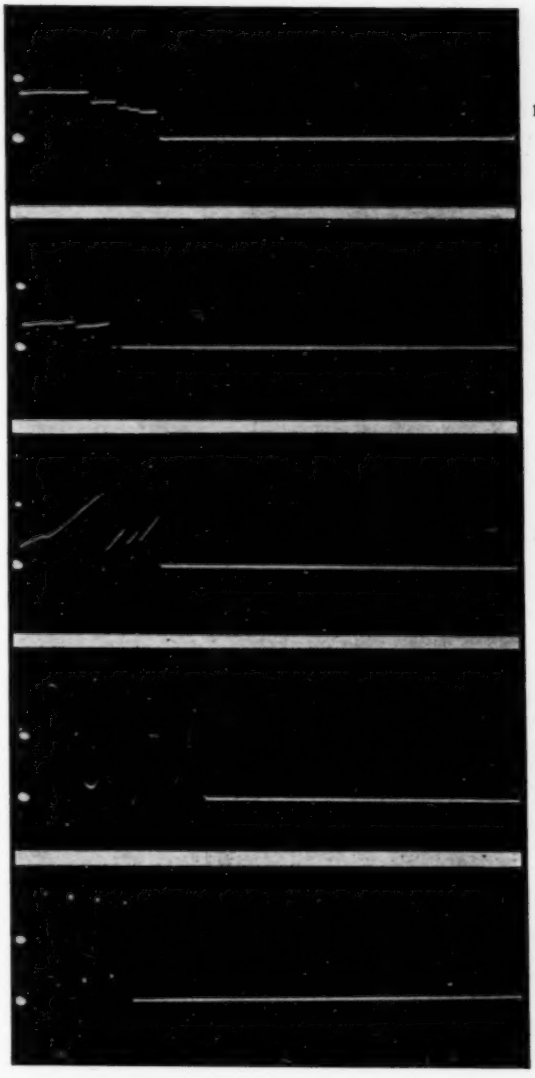


FIG. 3. Potential between electrodes of inactive silver as they are brought together to discharge a condenser through various inductances.

than for carbon, and for $L=0.01$h it is clear that conditions are more nearly adequate to establish the arc for active silver than for carbon, both of which observations are in line with the above figures. In the two lowest traces of Fig. 2 the arc is extinguished before the energy is consumed, presumably when the arc current has decreased to 0.03 amp; and then finally during the time interval just before contact closure there is a new series of discharges of the oscilloscope plate capacity.

ELECTRODE SEPARATION AT FIRST DISCHARGE

From examination of a large number of oscilloscope traces it is found that, although the elapsed time from the first partial discharge of the oscilloscope plates to final closure varies considerably, this time seems to be unrelated to the inductance and to be consistently higher for active silver and carbon than for inactive silver. For active silver and carbon the time is, on the average, about $t'=15\times10^{-6}$ sec and for inactive silver $t'=4\times10^{-6}$ sec. The displacements of the moving electrode in these times are, respectively, 6000A and 1600A. If one could be sure that the energy of the discharge were in each case insufficient to roughen or smooth the electrode surfaces appreciably, these distances would be the actual separations of the surfaces when the first partial discharge of the oscilloscope plate capacity takes place.[3] The nominal electric field between the electrodes at the first discharge would then be obtained by dividing the potential difference by these distances, giving 0.58×10^6 v/cm, 0.67×10^6 v/cm, and 2.2×10^6 v/cm respectively for the field between active silver surfaces, carbon surfaces and inactive silver surfaces. These values are written down in Table I.

TABLE I. Data for the first arc, average values.

	Active silver	Carbon	Inactive silver
Potential difference	35 volts	40 volts	35 volts
Time to closure	15×10^{-6} sec	15×10^{-6} sec	4×10^{-6} sec
Calculated separation	6000A	6000A	1600A
Nominal field	0.58×10^6 v/cm	0.67×10^6 v/cm	2.2×10^6 v/cm
Min. arc current (reference 1)	0.03 amp	0.03 amp	0.6 amp

These nominal values of electric field at which there is a partial discharge of the oscilloscope plate capacity between electrodes which are approaching each other can be compared with earlier measurements of the field at which an arc is struck between stationary electrodes. In the earlier tests measurements were made of the separation at which platinum electrodes discharged various low capacities in a very low inductance circuit. For 48 volts it was found[4] that the average nominal field was 0.60×10^6 volts/cm, which agrees well with the present measurements for active silver but is considerably lower than the field for inactive silver. It has been reported[2] that when electrodes which have not been specially cleaned are operated repeatedly, there is sufficient grease or other contamination on the surface so that on the first few closures they are in the active condition. It seems altogether probable that the surfaces of the platinum electrodes of the earlier experiments, at which only a very few closure arcs took place, were in the active condition at most closures and that conditions for arcing at active platinum surfaces are much the same as for active silver. In the platinum tests considerably greater spread in the nominal field was found than in the present tests with active silver electrodes; part of this difference may be due to some closures of the platinum surfaces at which they had become inactive, resulting in much higher fields, and part to occasional greater roughness of the platinum surfaces, due to the fact that there was no continual pounding together of the platinum wires as in the tests upon silver wires reported here.

THE "SMALL" ARCS

It has been one of the purposes of the preceding paragraphs to present a gradual approach toward the conclusion that the partial discharges of the oscilloscope

[3] For silver the energy in the oscilloscope plate capacity is $C'V_0^2/2=0.25$ erg, and one sees from the traces of Figs. 2 and 3 that somewhat more than half of this on the average is dissipated at the first partial discharge, perhaps 0.15 erg. The roughening produced by this small amount of energy can be calculated from the results of earlier work and turns out actually not to be negligible. In this earlier work measurements were made of the maximum separation at which electrodes are usually welded together by a discharge [reference 1, pp. 1091–1092, Figs. 7 and 8 and Eq. (3)]. For $V_0=48$ volts it was found that this critical separation is $s^*=0.05C^1$ cm. From the straightforward way in which the data were interpreted it is clear that for any amount of energy E, not necessarily that of a condenser charged to 48 volts, the general expression is $s^*=BE^1$ cm, where $B=2.2\times10^{-5}$ if E is expressed in ergs. For the energy of 0.15 erg this becomes $s^*=1200$A. If one neglects the fact that this determination of B was made for platinum and the present measurements are upon silver, this means that at the first partial discharge of the oscilloscope plate capacity the true separation of the electrodes was decreased by mounds of metal thrown up upon their surfaces to the extent of 1200A; and it seems to imply that the true separation of the electrodes at which this discharge occurred was greater than the calculated value (of 6000A or 1600A) by this amount. That there may be some flaw in this reasoning is apparent from the observation reported above that there is no dependence of t' upon circuit inductance; if the reasoning were correct, and the approaching electrode surfaces were perfectly smooth, one might expect that the many partial discharges which occur for low inductances would build up one mound on top of another and thus shorten t' considerably, so that observed values of t' would be greater for large inductances than for small, which is contrary to experience. The only sensible interpretation of the anomaly seems to be that successive discharges occur at different places on the electrode surface, no doubt at sharp points located in different places, so that the small mounds each of the order of 10^3A high do not pile on top of each other. It seems probable that the 1200A correction for one single partial discharge of the 40×10^{-12}-f capacity should be applied to each of the measured values of separation, making them 7200A and 2800A, respectively for active silver and for inactive silver. The corresponding nominal electric fields between the electrodes at the first partial discharge of the oscilloscope plates are 0.49×10^6 volts/cm for active silver, and 1.2×10^6 volts/cm for inactive silver. The correction resulting from this calculation is, however, so uncertain that it has not been incorporated into Table I nor used later in this paper.

[4] Reference 1, Table III, column 10.

plate capacity represented by the abrupt drops in the traces of Figs. 2 and 3 are really arcs which are fundamentally like those, for example, which have been discussed in connection with Figs. 1B and 1C of reference 2. Before this interpretation can be accepted, some of the peculiarities of these partial discharges must be discussed.

A natural point for first consideration is the absence of any direct record of the arcs on the traces reported here. This is, of course, accounted for by noting that the arc time should be about $\pi(L'C')^{\frac{1}{2}}=0.009\times10^{-6}$ sec, which is much too short to be seen.

A second point, which is not dismissed so abruptly and the interpretation of which may well lead to fundamental insight into the phenomena at the beginning of a short arc, is the fact that the open-circuit voltages following these arcs are always above zero, and often even above the known value of the arc voltage (reference 2, Table II). In our previous experience the open-circuit potential following an arc has been below the arc voltage and has been related in just the way one expects to the initial voltage, the arc voltage, and the circuit parameters.[5] It is natural to assume that this well-established relationship[6] still obtains and that the open-circuit voltages are now exceptionally high because the arc voltage, which is not directly observable on account of the very short duration of the arc, is itself exceptionally high. An unusually high arc potential at the very beginning of an arc seems reasonable enough, and for an arc of sufficiently short lifetime it is to be presumed that such an unusual potential can be observed through its consequences as in these experiments. There are, however, other possible explanations that must be ruled out before one can deduce a high and variable arc potential from the observation of high and variable open-circuit potentials following an arc.

More prosaic possible causes of high open-circuit potentials following an arc are (1) considerable resistance in the oscilloscope plate circuit so that the circuit is perhaps no longer mainly inductive, (2) early failure of the arc when the current is still considerable, and (3) partial recharge of the oscilloscope plate capacity from distributed capacity in the large inductance. Some simple considerations and experiments seem to rule out all of these possible explanations of the high open-circuit potentials.

The resistance which would be required in the oscilloscope circuit to make it critically damped is $(4L'/C')^{\frac{1}{2}}=140$ ohms. The dc resistance of this circuit is negligible, and it seems that even the high frequency resistance must be smaller than this value by a factor of almost 10^2, so no appreciable effect of the circuit resistance can be looked for.

In regard to early arc failure because the largest arc current may be only slightly above the minimum arc current for the electrodes, one finds that for silver electrodes the calculated largest arc current in one of the small arcs discharging the oscilloscope plate capacity is $(V_0-v)(C'/L')^{\frac{1}{2}}=0.34$ amp, which is a factor of 10 above the minimum arc current for active silver. The open-circuit voltage which would be observed for an arc ending when the minimum arc current of i' amp is reached is $V_1'=v-[(V_0-v)^2-Li'^2/C]^{\frac{1}{2}}$, which for active silver (with $i'=0.03$ amp) differs from $2v-V_0$ by less than 0.1 volt. It then seems that significantly premature arc failure from this cause cannot be expected for active silver. A conclusive experiment was performed by placing an inductance of 10×10^{-6}h in the lead from one of the electrodes to the oscilloscope plate. The presence of the inductance must decrease the largest arc current; and, if the arc voltage v were constant, the open-circuit voltage observed, after the arc fails when the minimum arc current i' is reached, would be increased in accord with the above formula; if this were the dominant effect of the inductance, the open-circuit voltage would be raised. The experimental observation is that this inductance *lowers* the open-circuit voltages of traces like those of Fig. 2, in many cases to values well below zero. This is attributable to increase in the calculated value of the duration of the arc from 0.009×10^{-6} sec to 0.06×10^{-6} sec and the assumption that the arc voltage v is initially exceptionally high but has reached its normal steady value in a time considerably less than 0.06×10^{-6} sec. A further observation bearing on this matter is the fact that open-circuit potentials for inactive silver (Fig. 3) are not significantly higher than for active silver (except for the 12-h trace); yet the maximum current in the arcs preceding these opens never reaches even to 0.6 amp, which has been previously determined as the *minimum* value for inactive silver; the situation regarding the arcs at inactive silver electrodes is thus not clear. It is, however, readily understandable that there should be many closures without an arc at all in the oscilloscope plate circuit (by actual count, about half of all closures). But one must note that a trace without feature, like that of Fig. 1A of reference 2, may as well as not represent not a closure without a single discharge of the oscilloscope plates but rather a weld of the electrodes by the first such discharge which takes place.

The third possibility of partial recharge of the oscilloscope plate capacity from distributed capacity in the large inductance is not ruled out in all cases. The effective distributed capacity is of appreciable size and under some conditions it considerably increases the open-circuit voltage which is observed. For example, the initial open-circuit potential for the discharge of inactive silver electrodes through an inductance of 12h, on the top trace of Fig. 3, is consistently higher in a large number of experiments than are other opens. If a 10^6-ohm carbon resistor of small physical size is placed in the circuit between the 12-h coil and the electrode to which it is attached, this difference is no longer ob-

[5] See, for example, Figs. 11, 14, and 15 of reference 1, and Fig. 1B of reference 2.

[6] Equations (4) and (5) of reference 1.

served; the open for inactive silver is on the average about the same as all other opens. It seems clear from this experiment that the effective distributed capacity of the 12-h coil is comparable with C', and that it is sufficiently isolated by inductance in the coil itself so that it is not discharged with the discharge of C' at inactive silver contacts, but is so discharged at electrodes of active silver. A check on the latter hypothesis is obtained by experiments upon active silver in which carbon resistors having values between 10^3 and 10^4 ohms have been used to isolate the 12-h inductance. The open is then much higher, obviously because of partial recharge of C' after the little arc is over with the recharge having a time constant RC' between 0.04 and 0.4×10^{-6} sec, which is too short for the recharge to be observed; the recharge can be seen with a resistance of 10^5 ohms.

These tiresome considerations have been necessary in order to arrive, by a process of elimination, at the conclusion that the effective arc voltage is above normal in the small arcs which discharge the oscilloscope plate capacity. The arc voltage is, however, not above normal on the average when for active silver electrodes, in the experiment described above, the arc time has been increased by inserting a 10×10^{-6}-h inductance. In different words, one can say that the average arc voltage is high when the arc time is 0.009×10^{-6} sec, but not high when the time is 0.06×10^{-6} sec; the time required for the arc voltage to reach its "normal" steady value must lie in this range and is perhaps of the order of 0.02×10^{-6} sec. The transit time of a silver ion across a gap of 6000A with a potential difference of 20 volts is 2.0×10^{-10} sec, and the time for an uncharged atom to traverse this distance with a speed corresponding to the boiling point of silver is 8.0×10^{-10} sec. Thus, the time required for an arc to reach its steady state is of the order of 100 times the transit time of a silver ion or 25 times the transit time of an uncharged silver atom at the temperature of the boiling point. This roundabout conclusion leads to the idea that one should seriously attempt to study by direct observation the arc potential during the first 0.02×10^{-6} sec. Equipment for such tests is now being prepared.

FIELD EMISSION CURRENT

One notes that the traces of Fig. 2 end with lines sloping downward toward the right which are more or less straight for an appreciable distance before a final abrupt fall of the voltage to zero. In marked contrast, the potential across clean silver contacts (Fig. 3) never decreases gradually; the final abrupt potential fall is not preceded by the sloping line which is seen on each trace recording the closure of contacts of active silver or carbon.

The sloping lines for active silver or carbon represent a decrease of potential across the oscilloscope plate capacity at a uniform rate, and to a constant current from this capacity, a 45° slope corresponding to 1.3×10^7 volts/sec and to a current of 5.2×10^{-4} amp. The total current flowing between the contacts is greater than that calculated from the slope of the curve and the oscilloscope plate capacity by the amount of the current from the larger capacity C, which is quite small in some cases but large in others. If we neglect this latter current, the final current between the contacts just before closure is of the order of 10^{-3} amp. The abrupt closure takes place when the potential across the electrodes reaches a critical value V_a which is of the order of 8 or 10 volts. One can speculate that the current, which is 10^{-3} amp or more, is field emission current across a small gap between the electrodes, or is current through a resistance film the resistance of which decreases by a factor of about 3 along the straight sloping line reaching a final value of 10^4 ohms or less. On the assumption that we are observing field emission current some simple calculations can be carried through which correlate a number of observations and thus lend support to this interpretation.

Theoretical considerations not yet published have led to the hypothesis that the final sudden closure may be due to electrostatic forces between the surfaces. Just before this closure the electrodes are approaching each other at $w=4.0$ cm/sec; and in a typical average case the voltage between them is decreasing at a uniform rate $dV/dt=2.5\times10^7$ volts/sec, and a constant current $I_a=10^{-3}$ amp is flowing. If S_a is the separation at which sudden closure occurs, F_a the electric field during the time the current is constant, and t_a the time between actual closure and when closure would have occurred if there had been no acceleration by electrostatic forces, one finds $t_a=V_aC'/I_a=0.4\times10^{-6}$ sec, $S_a=wt_a=1.6\times10^{-6}$ cm, and $F_a=V_a/S_a=6\times10^6$ volts/cm. The steady current of 10^{-3} amp is flowing from one electrode to the other while the field between them is remaining constant at 6×10^6 volts/cm, and final sudden closure occurs from a separation of 160A. Although this calculated field is a little too small to give much field emission current, it is large enough to imply very strongly that we are really observing field emission currents just before the closure of electrodes of *active silver* or of *carbon*; the true field is presumed to be $\theta(6\times10^6)$ volts/cm, where θ is a factor introduced by the surface roughness. It is evident from the traces of Fig. 3 that there is no detectable field emission current at the closure of electrodes of *inactive silver*.

The assumption that the sloping lines at the closure of contacts of active silver or of carbon represent field emission current, and not current flowing through a film of foreign material of variable resistance, can be extended to interpret other features of the oscilloscope traces. From the field emission interpretation of the sloping lines one naturally comes to the presumption that the first partial discharge of the oscilloscope plate capacity is initiated by field emission electrons. The nominal field at this first partial discharge is much smaller than that which gives rise to the current I_a, and

the field emission current which it produces must then be less than 10^{-3} amp by a factor which is simply enormous. It is to be presumed that the first partial discharge is initiated by a very small number of electrons, perhaps a single one.

After one has embraced the hypothesis that the lines of negative slope just before closure on Fig. 2 represent field emission current and that the first partial discharge of the oscilloscope plate capacity is initiated by field emission, one can make some interesting qualitative comparisons of some of the features of traces like those of Figs. 2 and 3. The fact that the first discharge occurs between surfaces of carbon or active silver at a much lower calculated field than between surfaces of inactive silver can be attributed either to a higher work function for inactive silver, or to a smoother surface, as a result of which the true electric field is not so much greater than the calculated field as it is for the carbon or active silver surfaces. This assumed higher work function for inactive silver, or smoother surface, then explains the observation that sudden closure occurs for inactive silver before the field emission current can be detected. These qualitative consistencies between the features of Figs. 2 and 3 can be made quantitative by the use of the Fowler-Nordheim equation, but the uncertainties are great.

I am indebted to Mr. J. L. Smith for carrying out most of the experiments described here, and to him and to Dr. W. L. Brown and Mr. F. E. Haworth for much valuable discussion.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 9 SEPTEMBER, 1951

X-Ray Absorption in a Crystal Set at the Bragg Angle

H. N. CAMPBELL
United States Rubber Company, Passaic, New Jersey
(Received March 14, 1951)

Whereas extinction effects would be expected to decrease the x-ray transmission through a crystal when it is set at a reflecting angle, it is shown experimentally that in many cases the transmission increases at the Bragg angle.

This is believed to be due to the fact that the decrease in the effective absorption coefficient overshadows the losses due to reflection.

A reconsideration of Darwin's presentation of extinction allowing for variable μ shows that the transmission may increase or decrease depending on the values of μ and p. Thus the earlier observations of Bragg and co-workers are reconciled with the observations reported here.

The importance, to crystal structure workers, of determining how μ varies with direction in periodic media is obvious.

INTRODUCTION

IN crystal structure determinations by means of x-ray diffraction it is customary to consider corrections to be applied to measured intensities which allow for the effect of absorption in the specimen. This absorption correction is complicated by an additional effect, called extinction, which is supposed to increase the effective absorption by removing additional energy from the incident beam through reflection.

The purpose of this paper is to present some data which indicate that this simple picture is inexact.

In studying the effects of extinction there are two cases to be considered. The first is the reflection case, where the incident beam enters and the reflected beam leaves the same face of the crystal. The reflecting planes are parallel to this face, and the crystal usually is so thick that the transmitted beam is absorbed and does not emerge. Second, there is the transmission case, where the incident beam enters one face of the crystal and both transmitted and reflected beams leave a second parallel face. If the reflection plane is perpendicular to these faces, the total path length for each beam is the same. We are concerned with the transmission case, since only here can the transmitted beam be measured conveniently.

According to Darwin's theory,[1] extinction, whether primary or secondary, acts to increase the effective linear absorption coefficient. Consequently, when a beam of x-rays of unit intensity falls on a crystal of thickness Z', we should have a transmitted beam,

$$T_0=\exp(-\mu_0 Z),$$

when away from the Bragg angle and a transmitted beam,

$$T=\exp[-(\mu_0+g_1Q)Z],$$

and a reflected beam,

$$R=ZQ\exp[-(\mu_0+g_2Q)Z],$$

when the beam is at the Bragg angle. Here $Z=Z'\sec\theta$, μ_0 is the ordinary linear absorption coefficient, Q is the structure factor, and g_1 and g_2 are extinction coefficients which may be equal.

The important consequence of Darwin's theory is that if we consider the ratios,

$$T/T_0=\exp(-g_1QZ) \quad \text{and} \quad R/T_0=ZQ\exp(-g_2QZ),$$

[1] C. G. Darwin, Phil. Mag. 43, 800 (1922).

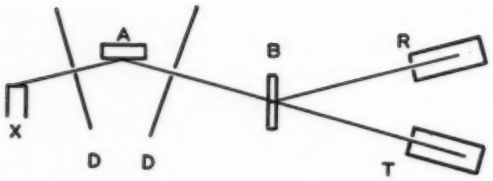


FIG. 1. Experimental arrangement. *X*, cathode of x-ray tube; *A*, calcite monochromater crystal; *B*, specimen crystal; *T*, Geiger counter to measure transmitted beam; *R*, Geiger counter to measure reflected beam; *D*, limiting apertures.

it is evident that neither the transmitted beam nor reflected beam at the Bragg angle can be greater than the beam transmitted off the Bragg angle.

Experimental

The transmission case was investigated to determine if the expected increase in effective absorption could be observed. This problem had previously been considered by W. L. Bragg and co-workers as far back as 1921 with results which will be discussed below.

The experimental arrangement is shown in Fig. 1. A beam of copper x-rays was rendered parallel and monochromatic by reflection from a cleavage face of a calcite crystal "*A*." The cross section of the beam was about 0.5 mm×5 mm. The angular divergence of this beam was less than 15 seconds, and in some of the work second-order short wavelength x-rays were eliminated by dropping the potential on the x-ray tube below 16 killovolts. Some of the experiments were carried out using copper K_β, since the K_α doublet probably was not resolved. This beam was allowed to fall on the crystal under investigation, "*B*," which was a thin slice about 1 mm thick and 1 cm square cut so that the reflection plane was perpendicular to a large face. Two Geiger counters were set up, "*T*" to catch the transmitted beam and "*R*" to catch the reflected beam. The crystal was turned through the Bragg angle, and intensity as a function of angle was determined for both beams.

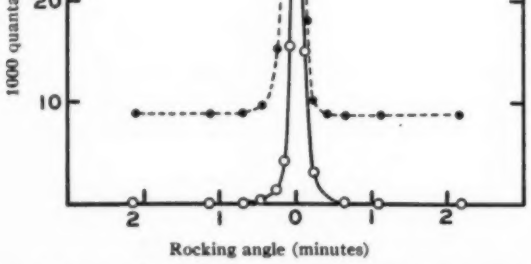


FIG. 2. Calcite ($10\bar{1}$, 1) reflection. Etched crystal *ca* 1 mm thick. λ=1.54A+0.77A.

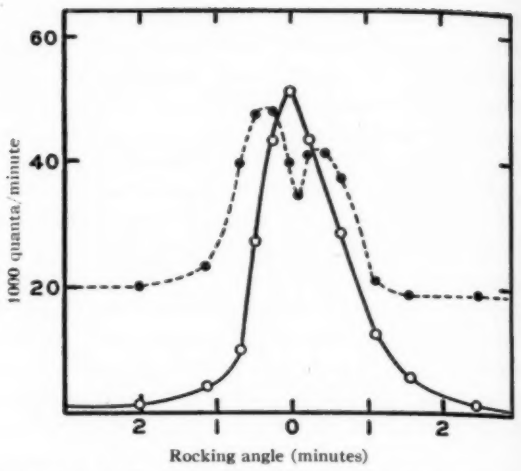


FIG. 3. Calcite ($10\bar{1}$, 1) reflection. Ground crystal *ca* 0.4 mm thick. λ=1.54A+0.77A.

Figures 2, 3, 4, and 5 show how the intensity varies for four different crystals. The solid line represents the diffracted beam; the broken line is the transmitted direct beam. In these experiments there was some second-order short wavelength x-rays present.

Instead of the transmitted beam decreasing as expected, it actually increases in most cases. The effect of difference in crystal perfection is definite.

Figure 2 is a calcite crystal. It was cut about 3 mm thick and reduced to about 1 mm by etching with hydrochloric acid. This crystal is quite perfect as judged from the half-width of the curve. The increase in the transmitted beam is striking.

Figure 3 is a similar calcite crystal but prepared by grinding down a thicker slice on emery paper. This grinding introduced imperfections in the crystal as judged by the increased half-width of the reflection. The relative increase in intensity of the transmitted beam is considerably less than for the more perfect crystal.

Figure 4 is a lepidolite mica. This is a relatively weak

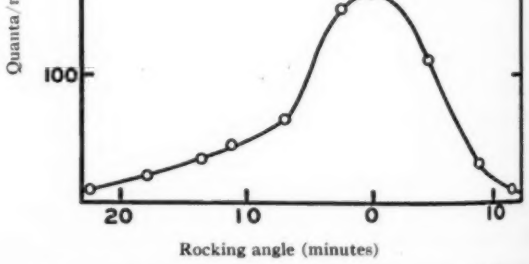


FIG. 4. Lepidolite mica, weak reflection at $2\theta=32.5°$. λ=1.39A+0.69A.

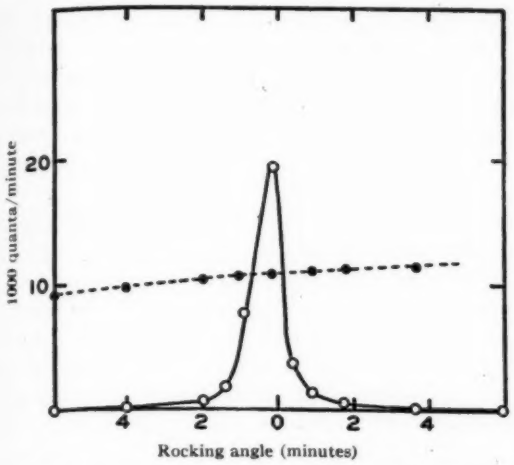


FIG. 5. Hexamethylenetetramine (100) reflection. Crystal *ca* 2 mm thick. $\lambda = 1.39A + 0.69A$.

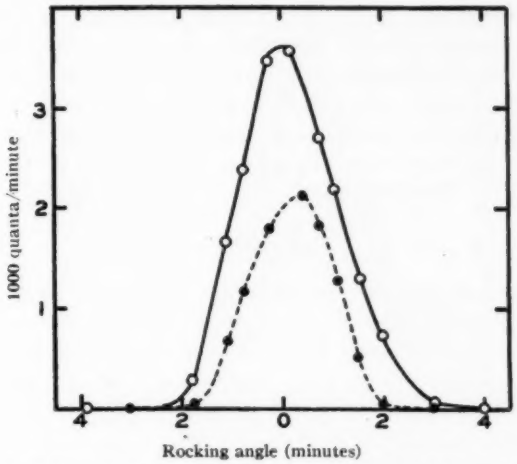


FIG. 7. Calcite ($10\bar{1}$, 1) reflection. Ground crystal 0.75 mm thick. $\lambda = 1.54A$. Incident beam 1.2×10^6 quanta/min.

reflection, and the crystal is quite imperfect. The increase in the transmitted beam is just perceptible.

Figure 5 is a rather poor crystal of hexamethylenetetramine. There is no increase in the transmitted beam but there is definitely no decrease.

In all cases except the weak lepidolite reflection, the reflected intensity is greater than that transmitted by the crystal when off the Bragg angle.

Figures 6 and 7 show the effect in calcite crystals when the second-order short x-rays are eliminated by dropping the potential on the x-ray tube.

Figure 6 is an etched crystal 0.40 mm thick. The incident intensity was 1.2×10^6 quanta per minute. The intensity transmitted off the Bragg angle was 1500 quanta/min. This corresponds to an absorption coefficient of 167.0 cm^{-1}. At the Bragg angle the reflected intensity is 36,700 quanta per minute and the transmitted intensity is 33,900 quanta per minute, corresponding to an effective absorption coefficient of 89.0 cm^{-1}. The calculated absorption coefficient is 192 cm^{-1}.

Figure 7 is a ground calcite crystal 0.75 mm thick. It showed practically no transmission off the Bragg angle. The calculated transmission (assuming an absorption coefficient of 192 cm^{-1}) should have been about 1 quantum/min, which was not detectable above the cosmic ray background of 73/min. This crystal, however, transmits 2100 quanta per minute at the position of peak reflection, giving an effective absorption coefficient of 85 cm^{-1}. The relative displacements of the peaks is due to slight misalignment of the crystal.

The effect is also shown in Fig. 8 for this same crystal. This is a 30-minute exposure to the direct and reflected beams, on a film mounted between the crystal and the Geiger counters. When the crystal is set off the Bragg angle, the intensity is too low to register in this length of time. When the crystal is set at the Bragg angle, both beams register at about the same intensity. Close inspection of this picture shows that the crystal is not uniform and that the intense regions of both beams come from the same parts of the crystal.

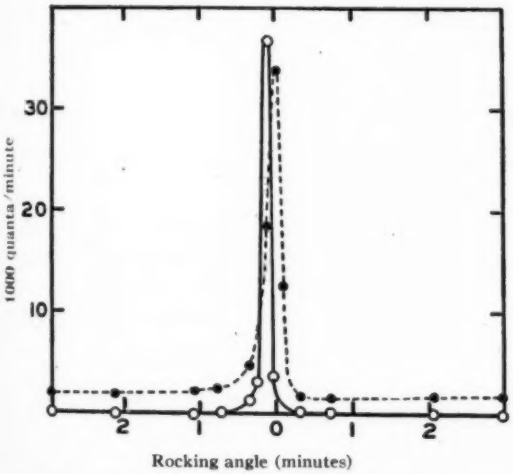


FIG. 6. Calcite ($10\bar{1}$, 1) reflection. Etched crystal 0.40 mm thick. $\lambda = 1.54A$. Incident beam 1.2×10^6 quanta/min.

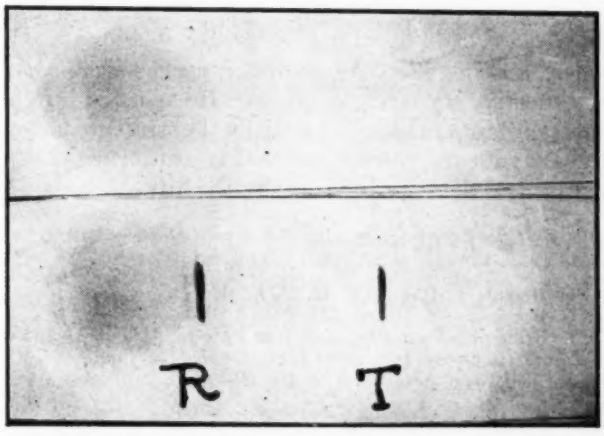


FIG. 8. Thirty-minute exposure to both beams from the same crystal as Fig. 7. Lower picture at the Bragg angle. Upper picture about 5 minutes off the Bragg angle. $\lambda = 1,54A$.

Borrmann[2,3] has reported the same effect with quartz and calcite crystals. He uses a technique similar to Lonsdale's divergent beam method. He finds that if a divergent beam falls on a crystal, some of the rays will be going in a direction to cause a Bragg reflection. He then finds superimposed on the general background intensified lines due to both the reflected and transmitted beams.

DISCUSSION

The only possible explanation of these results is that at the Bragg angle there is an over-all decrease in the macroscopic effective linear absorption coefficient. This does not necessarily mean that the absorbing power of individual atoms depends on direction; but only that the periodic structure shows preferred directions where absorption is a minimum.

Similar phenomena are known to exist in the case of wave propagation in periodic media[4] and in the case of multiple parasitic radio antenna arrays, but in neither case has the theory been extensively treated for three dimensions.

Von Laue[5] has considered a possible explanation of Borrmann's results based on the use of complex absorption coefficients. However, his predicted intensity distributions are not the same as observed in this work. Ewald has suggested a mechanism based on standing waves, which is the three-dimension analog of the coaxial cable transmission line problem.

However, it is not the intention here to discuss why the absorption decreases, but to point out some of its consequences.

When a beam falls on a crystal at the Bragg angle, there results a transmitted beam and a reflected beam. Both of these are subject to absorption, and each scatters in the direction of the other.

Darwin[1] has shown that this results in a transmitted beam,

$$T=\tfrac{1}{2}\exp(-\mu z)+\exp[-(\mu+2p)z],$$

and a reflected beam,

$$R=\tfrac{1}{2}\exp(-\mu z)-\exp-[(\mu+2p)z],$$

where μ is the effective absorption coefficient and p is the effective scattering coefficient. He assumed μ to be constant but p (which is essentially the structure factor) to vary with angle. If we realize that μ also varies but is always less than μ_0, and consider the ratio,

$$T/T_0=\tfrac{1}{2}\exp[(\mu_0-\mu)z]+\tfrac{1}{2}\exp[(\mu_0-\mu-2p)z],$$

we see that if $(\mu_0-\mu)z$ is large enough so that $\exp[(\mu_0-\mu)z]$ is greater than 2, then T must be greater than T_0. At the same time R can become greater than T_0 if p is large enough. However, if μ_0 is so small that $\exp(-\mu_0 z)$ is more than 50 percent, and at the same time p is large enough so that the second term is very small, then T may be less than T_0.

That is to say, a decrease in intensity could be expected only when μ_0 is very small and p very large. Even then, T/T_0 cannot be less than $\frac{1}{2}$.

Bragg, Bosanquet, and James[6] tried to measure primary extinction by essentially the method under consideration. Using a NaCl crystal and short wavelength radiation, they found about 20 percent decrease in the transmitted beam. This was considerably less than anticipated, and they concluded that this was not a satisfactory method of determining primary extinction. Complete details were not given; but μ_0 is small and p large, so that a small decrease in T might be expected in this case and their results are, therefore, in accord with this treatment.

On differentiating the expression for R, we find a maximum value of R when

$$Z_m=(1/2p)\ \log[(\mu+2p)/\mu]=(1/p)\ \tanh^{-1}[p/(\mu+p)].$$

Thus Z_m will vary with the intensity of reflection and be smaller for intense reflections, as observed by Bragg, Bosanquet, and James; however, Z_m is a measure of p rather than a measure of secondary extinction as they assumed.

CONCLUSIONS

The decrease in effective absorption in a crystal at the Bragg angle has now been established by two separate methods on some 4 different crystals.

Apart from any quantitive aspects, the fact that, at the Bragg angle, the intensities of both transmitted and reflected beams can be greater than T_0 shows that extinction as usually visualized cannot be active, but that some opposite effect is operating.

The experiments of Bragg *et al.* cannot be considered as supporting evidence for the existence of extinction, since they are in complete agreement with the present treatment.

The effects reported are only indirectly dependent upon crystal perfection or imperfection. The effect depends only on the fact that both μ and p are functions of direction in a periodic medium. The manner in which p depends upon direction, crystal symmetry and perfection, and atomic structure has long been the concern of physicists interested in crystal structure. It is now necessary to determine how μ is affected by these same variables.

[2] G. Borrmann, Physik. Z. **42**, 157 (1941).
[3] G. Borrmann, Z. Physik **127**, 297 (1950).
[4] L. Brillouin, *Wave Propagation in Periodic Media* (McGraw-Hill Book Company, Inc., New York, 1946).
[5] M. Von Laue, Acta Cryst. **2**, 106 (1949).
[6] Bragg, Bosanquet, and James, Phil. Mag. **42**, 1 (1921).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 9 SEPTEMBER, 1951

On the Theory of Random Noise. Phenomenological Models. I*

DAVID MIDDLETON
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received March 30, 1951)

Three phenomenological models are considered from which can be constructed a macroscopic statistical description of varieties of electronic noise, of which shot, thermal, and Barkhausen noise, electron multiplier and precipitation noise, clutter, ignition, and impulsive random noise in general are representative examples. The models examined are (I) *non-overlapping, periodic noise waves*, common in pulse-time modulation and other communication schemes, where the amplitude, phase, duration, and epoch in a period interval are subject to statistical variations; (II), *non-overlapping, nonperiodic disturbances*, encountered in servomechanism operation and keyed-carrier communication techniques, for instance, which are like (I), but lack the basic periodic structure; and (III), *poisson noise*, consisting of the superposition of independent, randomly occurring elementary impulses. Much of the electronic noise mentioned above belongs to this more comprehensive type, where overlapping of the basic pulses is the characteristic feature. Because *all* (second-order) moments are required in general for the analysis of noise in nonlinear systems, the attempt is made here to determine explicitly on the basis of the appropriate model the second-order probability density W_2 in the important stationary cases. For noise of types (I) and (II) this appears impractical except in the simplest cases: only the lower order moments prove tractable. However, for poisson noise (III) an explicit treatment is possible for impulsive random noise, nearly normal random noise, and for the limiting, normal random cases (of which shot and thermal noise are examples). In Part I the main features of the models (I–III) are discussed, and the general probability density $W_1(X_1, t_1; \cdots; X_s, t_s)$ in the nonstationary instances is formally constructed. In Part II, the distribution density for nearly normal random noise is given, the first and second (second-order) moments of the various distributions are determined, and from these in turn are found the spectral distribution of the energy in the random waves.

1. INTRODUCTION

A COMPLETE statistical description of electronic random noise can be given with the aid of the theory of time-dependent fluctuation phenomena. In its details such a description must depend ultimately on a microscopic, classical, or quantum-mechanical treatment, since the basic origin of the noise lies in the discrete structure of matter; but for the present purpose of providing a large-scale statistical description, general phenomenological models are sufficient. These models are satisfactory not only for electronic noise, but for other noise and fluctuation mechanisms as well. (A discussion of detailed structures is reserved for a later paper.) The Brownian movement,[1,2] diffusion in gases,[1] concentration fluctuations in colloidal suspensions,[1] and of chief interest here, noise currents and voltages in linear and nonlinear systems[3–8] are important physical examples. To date, however, the analytical description of noise systems has been largely limited to those cases where the random noise belongs to the so-called *fluctuation types*,[9] which in turn belong to a normal random process whose statistical properties are well known.[10] Brownian movement and the electronic cases of shot and thermal noise[10a] are representative. In the present paper, on the other hand, we are concerned with the equally important situations involving more general types of impulsive interference. These it is convenient to discuss in three categories:

(I) *Periodic, non-overlapping impulse noise:* This is a case where the disturbance consists of a train of pulses, random in amplitude, duration, and phase, and which occur at random intervals during each repetition period. Such waves are common in certain types of (pulse-position) modulation systems, in electronic computing schemes, in the electronic calculation of correlation functions,[11] and when normally periodic systems are "jittered" by noise.

(II) *Nonperiodic, non-overlapping impulsive noise:* This type is quite similar to (I), but with the essential difference that the durations of the periods may be random. Wave trains of this kind are encountered in the operation of servomechanisms[12] in certain

* The research reported in this document was made possible through support extended Cruft Laboratory, Harvard University, jointly by the Navy Department (ONR), the Signal Corps, U. S. Army, and the U. S. Air Force, under Contract N5ori-76 T. O. I.

[1] S. Chandrasekhar, Revs. Modern Phys. **15**, 1 (1943).

[2] M. C. Wang and G. E. Uhlenbeck, Revs. Modern Phys. **17**, 323 (1945).

[3] W. R. Bennett, J. Am. Acoust. Soc. **15**, 165 (1944); Bell System Tech. J. **26**, 139 (1947).

[4] S. O. Rice, Bell System Tech. J. **23**, 282 (1944); **24**, 46 (1945); **27**, 109 (1948).

[5] A. Blanc-Lapierre, dissertation (University of Paris) (Masson et Cie, Paris, 1945).

[6] J. H. Van Vleck and D. Middleton, J. Appl. Phys. **17**, 940 (1946).

[7] D. Middleton, J. Appl. Phys. **17**, 778 (1946); Quart. Appl. Math. **5**, 445 (1948); J. Appl. Phys. **19**, 817 (1948); Proc. Inst. Radio Engrs. **36**, 1467 (1948); J. Appl. Phys. **20**, 334 (1949); Quart. Appl. Math. **7**, 128 (1949); *ibid.*, **8**, 59 (1950); J. Appl. Phys. **21**, 734 (1950).

[8] J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals* (McGraw-Hill Book Company, Inc., New York, 1950), M.I.T., Radiation Laboratory Series No. 24. References 3 to 8 list further papers dealing with the large-scale problems of noise in electrical communications; see also reference 2.

[9] A paper by M. Kac and A. J. F. Siegert [J. Appl. Phys. **18**, 383 (1947)] examines the exceptional case of *rectified* and filtered, originally normal random noise; owing to rectification (and filtering) the distribution law of noise is no longer gaussian.

[10] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, New Jersey, 1946).

[10a] For a discussion of thermal noise see S. A. Goudsmit and P. Weiss, "Statistics of circuit noise," M.I.T. Radiation Laboratory Report No. 192, January 29, 1943. See also references 2 and 8.

[11] Lee, Cheetham, and Wiesner, Proc. Inst. Radio Engrs. **38**, 1165 (1950).

[12] James, Nichols, and Phillips, *Theory of Servo-Mechanisms* (McGraw-Hill Book Company, Inc., New York, 1947), M.I.T. Radiation Laboratory Series No. 25, Chapter 6.

types of communication (keyed carrier transmission), and the model also provides information on the spectral distribution of pressure-broadened spectral lines.[13]

(III) *Poisson noise:* The distinctive feature of this model is the completely random occurrence of the elementary impulses; overlapping is common and characteristic. Familiar physical examples are atmospheric and solar "static," precipitation noise, ignition and lightning discharges, the interference effects of randomly oriented scatterers in the medium of propagation, etc. Unlike fluctuation noise or Brownian motion, where the density (in time) of the individual effects is so great that overlapping is extremely frequent, this latter type of disturbance is characterized by relatively few impulses per unit time, with a weak overlapping of individual effects, and so does not belong to a normal random process. Only in the limit of a high impulse density does the poisson noise (so named because of its statistical structure) exhibit gaussian properties (see Sec. 4).

Our purpose, therefore, is to construct and examine the general statistical models from which the properties of the principal noise types (I–III) may be derived. The detailed applications are reserved for a later paper, as is also the explicit study of the "atomic" structure of the noise. Earlier efforts were restricted mainly to a derivation of first and second moments of the random fluctuations of the poisson type (e.g., Campbell's theorem[14–17]), although more recently in some instances correlation functions and spectra for types I and II have also been obtained.[8,12] In the case of poisson noise a simpler and less general model has been studied by Rice,[4] Blanc-Lapierre,[5] and Boonimovich,[18] in their discussion of the shot effect. Their results are applicable to a limiting case of fluctuation noise; no treatment of the low density impulsive noise has been given, nor is the situation of intermediate amounts of overlapping among individual effects considered.

All the statistical properties of a random wave, namely, the moments, characteristic functions, spectra, and auto- and cross-correlation functions, as well as the probability distributions from which they are derived, may be determined once the particular stochastic process to which it belongs has been specified. This accordingly requires that one know the set of frequency functions $W_1, W_2, \cdots, W_n$ which completely describes the process.[2] It is assumed throughout that we are dealing with a continuous stochastic system, so that $W_n(n\geq 1)$ are, strictly speaking, probability densities. For most applications, however, only the first- and second-order densities W_1, W_2 are required, even when the process is not markoffian. These are sufficient to give us the quantities of chief macroscopic physical interest, which are

(*a*) *the average or steady component,*
(*b*) *the mean-square amplitude,* and
(*c*) *the correlation function,* from which one finds in turn by a theorem of Wiener[19] and Khintchine[20] *the mean intensity* spectrum of the disturbance, which indicates the distribution of the wave's energy as a function of frequency.

In the case of a number of random variables, $X, Y, Z, \cdots$, the desired first- and second-order probability densities have the following significance:

$$W_1(X, Y, Z, \cdots; t_1)dXdYdZ\cdots$$
= the joint probability that $X, Y, Z\cdots$ lie in the ranges $(X, X+dX)$, $(Y, Y+dY)$, $(Z, Z+dZ)$, $\cdots$ at some initial time t_1; (1.1)

$$W_2(X_1, Y_1, Z_1, \cdots, t_1; X_2, Y_2, Z_2, \cdots, t_2)\cdot dX_1dX_2dY_1dY_2dZ_1dZ_2\cdots$$
= the joint probability that $X, Y, Z\cdots$ simultaneously fall in the intervals (X_1, X_1+dX_1), (Y_1, Y_1+dY_1), (Z_1, Z_1+dZ_1), $\cdots$ at a time t_1, *and* in the ranges (X_2, X_2+dX_2), (Y_2, Y_2+dY_2), (Z_2, Z_2+dZ_2), $\cdots$ at a time t_2 later, (1.2)

with similar interpretations for the high order densities $W_n(n\geq 3)$. When the process is stationary, i.e., when the underlying physical mechanism does not change with time, one has instead to consider only the time interval $t=t_2-t_1$ between observations, since the choice of time origin is now arbitrary. Actually, once W_2 has been found,[21,22]† W_1 can readily be obtained from it on letting this time interval between the two sets of observations, designated by subscripts 1 and 2, respectively, become indefinitely large, i.e., $t=t_2-t_1\to\infty$, for then one has (in the stationary cases $t_1=0$)

$$\lim_{t\to\infty} W_2(X_1, Y_1, Z_1, \cdots; X_2, Y_2, Z_2, \cdots; t) = W_1(X_1, Y_1, Z_1\cdots)W_2(X_2, Y_2, Z_2\cdots), \quad (1.3)$$

[13] H. M. Foley, doctoral dissertation, University of Michigan, 1942; see also, reference 8, pp. 45, 46.
[14] N. Campbell, Proc. Cambridge Phil. Soc. **15**, 117, 310 (1909).
[15] E. N. Rowland, Proc. Cambridge Phil. Soc. **32**, 580 (1936); see also, J. H. Whittaker, *ibid.* **33**, 451 (1937).
[16] A. Khintchine, Bull. acad. sci. U.R.S.S. Sér. math., No. 3, 313 (1938).
[17] R. S. Rivlin, Phil. Mag. **36**, 688 (1945).
[18] V. Boonimovich, J. Phys. (U.S.S.R.) **10**, 35 (1946).
[19] N. Wiener, Acta Math. **55**, 117 (1930).
[20] A. Khintchine, Math. Ann. **109**, 604 (1934).
[21] N. M. Blachman, J. Appl. Phys. **20**, 38 (1949); also D. Middleton, J. Appl. Phys. **20**, 334 (1949); Quart. Appl. Math. **7**, 128 (1949); *ibid.*, **8**, 59 (1950).
[22] F. L. H. M. Stumpers, Proc. Inst. Radio Engrs. **36**, 1080 (1948).

† It should be noted that although one is frequently interested in a single random variable X, so that $W_2(X_1, X_2, t)$ is all that is needed, it is sometimes necessary to consider jointly a large number of random quantities. For example, in the study of narrow-band waves (references 3, 4, 6, and 10) where it is possible to express the random quantities in terms of an envelope R and phase θ, themselves random variables, slowly varying compared with the mean frequency component of the disturbance, W_2 must be known for R and θ at two arbitrary times. Accordingly, a four-dimensional distribution density $W_2(R_1, \theta_1; R_2, \theta_2; t)$ in the stationary case is needed. In the theory of the reception of frequency-modulated waves in noise (references 7, 21, and 22) an eight-dimensional distribution density is required, similar to $W_2(R_1, \theta_1; R_2, \theta_2; t)$, but for the envelope, phase, and their first time derivatives. Still other examples can be found in the literature (references 3, 4, and 7).

since there is no longer any "memory" of or correlation between events $(X_1, Y_1, Z_1\cdots)$ at the initial time (t_1) and events $(X_2, Y_2, Z_2\cdots)$ at the later time (t_2). We remark finally that for linear systems the macroscopic quantities (a)–(c) above are directly proportional to the appropriate first and second moments (considered at a given time or jointly at two distinct times). But for nonlinear systems (which are of central importance in communication systems) it is generally not enough to know these moments: the higher moments are also needed in calculating the mean, the mean square, the correlation functions, and associated spectra of the modified waves. Thus, the (second-order) density W_2 is required. Furthermore, in nonlinear cases it is the fourier transform (characteristic function[10]) of the probability density W_2 which proves more convenient analytically here, rather than the probability density itself, whose direct use in linear cases is more natural. The details of the method for a variety of applications are summarized in a recent paper.[23]

Before we proceed to general problems with more than one random variable, it is helpful to consider first a few simple examples, which may serve as a guide to the more comprehensive formulation. Let us treat, therefore, a single random variable X, which we further postulate to be the weighted sum of K elementary impulses $e(t-t_k')$ in a given interval of time $(0, T)$. Then we can write

$$X_K=\sum_k^K a_k e(t-t_k')=X(t; a, t')_K, \tag{1.4}$$

where a_k is an amplitude (or weighting factor) and t_k' is the epoch (relative to $t=0$) at which the typical kth pulse is initiated; X_K accordingly represents that part of the random wave X which occurs in the interval $(0, T)$. [For the models considered in this paper the impulses $e(t-t_k')$ are always of duration short compared with $(0, T)$.] The statistical character of the wave appears when we require the a_k or t_k' (or both) to be random variables, subject to a certain distribution law, say, $w(a, t')$. Precisely what this distribution law is, of course, depends on the physical process whose behavior we wish to describe.

Our task now is to determine the probability densities $W_1(X_1, t_1)$, $W_2(X_1, t_1; X_2, t_2)$, etc. For the moment, it is sufficient to illustrate the procedure by constructing the first-order density $W_1(X_1, t_1)$. Then, if we consider the ensemble of intervals $(0, T)$ containing all possible numbers of impulses $(K=0, 1, 2\cdots\infty)$, we can follow the method used by Rice[4] and others,[5,15–18] and write for the probability that at time $t_1(0\leq t_1\leq t)X$ lies in the range (X_1, X_1+dX_1)

$$W_1(X_1, t_1)_T dX_1=\sum_{K=0}^{\infty}\{\text{probability of exactly } K \text{ impulses in the interval } (0, T)\}\{\text{probability that if these are exactly } K \text{ impulses in } (0, T), X_1 \text{ lies in the range } (X_1, X_1+dX_1) \text{ at the time } t_1(0\leq t_1\leq T)\} \tag{1.5a}$$

$$=\sum_{K=0}^{\infty} W_1(K)_T P_1(K|X_1, t_1)_T dX_1. \tag{1.5b}$$

The conditional probability $P_1(K|X_1, t_1)_T$ is given by

$$P_1(K|X_1, t_1)_T=\int_{-\infty}^{\infty}(d\xi/2\pi)\times\exp(-iX_1\xi_1)F_1(\xi_1, t_1)_{T,K}, \tag{1.6}$$

when F_1 is the characteristic function for the random variable X_1 considered in the interval $(0, T)$. From the definition of the characteristic function one has

$$F_1(\xi_1, t_1)_{T,K}=\langle\exp(iX_1\xi_1)\rangle_{\text{stat av.}}=\int\int_{-\infty}^{\infty} dadt' w(a, t')\exp[iX_1(t_1; a, t')_K\xi_1]. \tag{1.7}$$

Our problem is now reduced to (i), determining a suitable distribution for the possibly random parameters a, t', which govern the elementary impulses, and (ii), specifying how the elementary impulses are "compounded" in the time $(0, T)$. By (ii) we can distinguish between the three types of noise (I–III) discussed in the introduction. This procedure is easily extended to more complicated wave forms: for example, where the elementary impulses are $a_k e(t-t_k'; \tau_k)$, τ_k being a random duration of this typical kth pulse. In a similar fashion more than one wave train can be handled; one has then to consider the distribution of X_1, X_2, Y_1, Y_2, $\cdots$ and so on. These generalizations are examined in Sec. 2, where the simple technique outlined above is appropriately modified to give us the desired higher order, time-dependent probability densities.

2. NOISE MODELS: PROBABILITY DENSITIES FOR GENERAL, NONSTATIONARY SYSTEMS

Let us now construct in a general way the hierarchy of probability densities W_l, $(l=1, 2\cdots, s)$, which completely describes the nonstationary random process to which the set of general (real)‡ functions of time $X_j(t)$, $(j=1, 2\cdots s)$ belong, when each is considered during an interval of time $(0, T)$. We indicate the random nature of the X_j by writing explicitly $X_j=X_j(t; a_1^{(j)}, a_2^{(j)}, \cdots)$, where the $a_\mu^{(j)}$, $(\mu=1, 2\cdots)$ are random variables with the joint distribution density $w_j(a_1^{(j)}, \cdots)$; these stochastic parameters are

[23] D. Middleton, "Noise and nonlinear communication problems," *Symposium on Applications of Autocorrelation Analysis to Physical Problems*, Woods Hole, Massachusetts, June 13, 1949 (ONR, May, 1950), p. 24.

‡ A complex function may be handled if we consider the real and imaginary parts as two new functions; thus if all s $X_j(t)$ are complex, a $2s$-dimensional density is generated.

to be interpreted as random amplitudes, phases, frequencies, pulse durations, etc., as required by the physical problem, which also determines the explicit form of X_j.

We desire the lth-order, s-dimensional joint probability $W_l(X_1, t_1; X_2, t_2; \cdots; X_s, t_s)_T dX_1 dX_2 \cdots dX_s$ that each of the X_j $(j=1, \cdots s;\ s \geq l \geq 1)$ of our nonstationary system jointly lie in the ranges (X_1, X_1+dX_1) at time t_1, in the interval (X_2, X_2+dX_2) at time t_2, $\cdots$, and in the range (X_s, X_s+dX_s) at time t_s, where $0 \leq t_1 \leq t_2 \cdots \leq t_j \cdots \leq t_s \leq T$). In general, the s-functions X_j $(j=1 \cdots s)$ may be distinct, but not necessarily independent.

At this point it is necessary to give a more precise structure to the $X_j (j=1 \cdots s)$; the simplest and most useful one is to consider each X_j as the resultant of exactly K "events" in the interval $(0, T)$, so that we may write more fully $X_j = X_j(t_{j'}; a_{1,0(j)}{}^{(j)} \cdots a_{1,k(j)}{}^{(j)} \cdots a_{1,K(j)}{}^{(j)};\ a_{2,0(j)}{}^{(j)} \cdots a_{2,K(j)}{}^{(j)};\ \cdots)$, where $j=1, 2 \cdots s$; $j'=1, 2 \cdots$, $l \leq s$; $k(j)$ denotes the kth event in the series $k=0, 1, 2, \cdots K$ for the jth function X_j, and we write $k(j)$ to distinguish between different values of k which may occur for different choices of j. The superscripts on the random parameters $a^{(j)}$ likewise distinguish between different possible statistical properties among the $j(=1, 2 \cdots s)$ resultant waves X_j. Here, the term "event" may signify the arrival of an electron at the plate of a tube, as in the shot effect; the collision between molecules of gas, as observed in the pressure-broadening of spectral lines;[13] the occurrence of a random pulse, as in certain types of pulse-trains occurring in electrical problems; or any other elementary, random effect which establishes the explicit structure of the X_j. Taking into account now all possible numbers of events $(K=0, 1, 2 \cdots \infty)$ that can occur in the time $(0, T)$, we may write

$$W_l(X_1, t_1; \cdots; X_s, t_s)_T dX_1 dX_2 \cdots dX_s$$
$$= \sum_{K=0}^{\infty} W_1(K)_T P_l(K | X_1, t_1; \cdots; X_s, t_s)_T \cdot dX_1 dX_2 \cdots dX_s$$
$$= \sum_{K=0}^{\infty} [\text{probability of exactly } K \text{ events in the interval } (0, T)] \times [\text{joint probability that if there are exactly } K \text{ events in } (0, T), X_1 \text{ lies in the range } (X_1, X_1+dX_1) \text{ at a time } t_1 (0 \leq t_1 \leq T), X_2 \text{ falls in the interval } (X_2, X_2+dX_2) \text{ at time } t_2 (t_1 \leq t_2 \leq T), \cdots \text{ etc.}]. \quad (2.1)$$

The conditional probability P_l is found from

$$P_l(K | X_1, t_1; \cdots; X_s, t_s)_T = \int_{-\infty}^{\infty} \frac{d\xi_1}{2\pi} \cdots \int_{-\infty}^{\infty} \frac{d\xi_s}{2\pi} \exp\{-i \sum_{j=1}^{s} X_j \xi_j\} \times F_l(\xi_1, t_1; \cdots \xi_s, t_s)_{T,K}, \quad (2.2)$$

where F_l is the characteristic function for the random variables $X_1 \cdots X_s$ considered in the interval $(0, T)$. By the definition of the characteristic function[10] one has

$$F_l(\xi_1, t_1; \cdots; \xi_s, t_s)_{T,K} = \{\exp(i \sum_{j=1}^{s} \xi_j X_j)\}_{\text{stat av.}} = \prod_{j=1}^{s} \int_{\mathbf{a}^{(j)}} w_j(\mathbf{a}^{(j)}) \mathbf{da}^{(j)} \exp[i \xi_j X_j(t_{j'}; \mathbf{a}^{(j)})], \quad (2.3)$$

where

$$w_j(\mathbf{a})^{(j)} \mathbf{da}^{(j)} = w_j(a_{10(j)}{}^{(j)}, \cdots a_{1k(j)}{}^{(j)} \cdots a_{1K(j)}{}^{(j)}; a_{20(j)}{}^{(j)} \cdots a_{2K(j)}{}^{(j)}; \cdots) da_{10(j)}{}^{(j)} da_{11(j)}{}^{(j)} \cdots,$$

and the integration in (2.3) is to be performed over all allowed values of the random parameters. Note that for any one strip of K events and any one member X_j of the set $(j=1, 2, \cdots s)$ the parameters $a_{1,k(j)}{}^{(j)}$, $a_{2,k(j)}{}^{(j)}$, etc., are governed by the same probability laws; however, these distributions (w_j) may differ for different members of the set, and it is also assumed here that there is no correlation between the random parameters of the set. Our formulation (2.1) is further limited by the assumption that the basic probability $W_1(K)_T$ for the occurrence of K events in $(0, T)$ is identical for the $j(=1, 2 \cdots s)$ members X_j of the set. All this, nevertheless, does not restrict the usefulness of the detailed models for the physical problems considered here. Moreover, it is usually not necessary to distinguish among the distributions w_j, so that $w_j = w$ for all j and the characteristic functions (2.3) may be reduced to

$$F_l(\xi_1, t_1; \cdots; \xi_s, t_s)_{T,K} = \int_{\mathbf{a}} w(\mathbf{a}) \mathbf{da} \exp(i \sum_{j}^{s} \xi_j X_j(t_{j'}; \mathbf{a})_K) \quad (2.4)$$

from the definition of the frequency functions w_j. When K is zero, i.e., when there are no events in the interval $(0, T)$, the X_j all vanish, the characteristic function becomes unity, and the conditional probability P_l is

$$P_l(0 | X_1, t_1; \cdots; X_s, t_s)_T = \prod_{j=1}^{s} \delta(X_j - 0), \quad (2.5)$$

where δ is the Dirac singular function.

The resultant (X_j) of exactly K events in $(0, T)$ is now taken to be a *linear superposition* of elementary impulses e_j. We distinguish two important cases: (a) broad-band disturbances, where the width of the energy spectrum of the elementary pulse is comparable to the central frequency of the spectrum, and (b) narrow-band waves, whose spectral energy spread is small relative to the central frequency of the band. Therefore, we can write explicitly for $(X_j)_K$ in these two instances:

$$X_j(t_j; \mathbf{a})_K = \sum_{k(j)=0}^{K} a_{k(j)} e_j(t_{j'} - t_{k(j)}{}'; r_{k(j)}) \cos\psi_{k(j)}, \quad (2.6a)$$

and

$$X_j(t_{j'};\mathbf{a})_K=\sum_{k(j)=0}^{K} a_{k(j)}e_j(t_{j'}-t_{k(j)}';r_{k(j)}) \times\cos[\omega_0(t_{j'}-t_{k(j)}')+\psi_{k(j)}]. \quad (2.6b)$$

Here the parameters $(\mathbf{a}=)a_{k(j)}$, $t_{k(j)}'$, $r_{k(j)}$, $\psi_{k(j)}$ represent the random amplitudes, times of occurrence, durations, and phases of the elementary pulses, respectively, and the statistical properties of the resultant $(X_j)_K$ then clearly depend not only on the model's structure as given by Eq. (2.6), but on the distribution $w(\mathbf{a})$ which governs the random parameters. It is also convenient in narrow-band cases to resolve the wave $(X_j)_K$ into slowly and (relatively) rapidly varying parts, and to examine the distribution laws which govern the slowly varying portion. Qne has then to consider instead of the single quantity X_j, the two auxiliary random variables Y_j and Z_j, obtained from Eq. (2.6b) by writing formally

$$X_j(t)_K=Y_j(t)_K\cos\omega_0 t+Z_j(t)_K\sin\omega_0 t =R_j(t)_K\cos[\omega_0 t-\theta_j(t)_K], \quad (2.7)$$

$$(R_j)_K=[(Y_j)_K{}^2+(Z_j)_K{}^2]^{\frac{1}{2}},\quad (\theta_j)_K=\tan^{-1}Z_j/Y_j\pm\pi q,$$

where

$$Y_j(t)_K=\sum_{k(j)=0}^{K} a_{k(j)}e_j(t-t_{k(j)}';r_{k(j)}) \times\cos[\omega_0 t_{k(j)}'-\psi_{k(j)}]=R_j(t)_K\cos\theta_j(t)_K. \quad (2.8a)$$

$$Z_j(t)_K=\sum_{k(j)=0}^{K} a_{k(j)}e_j(t-t_{k(j)}';r_{k(j)}) \times\sin[\omega_0 t_{k(j)}'-\psi_{k(j)}]=R_j(t)_K\sin\theta_j(t)_K; \quad (2.8b)$$

Y_j and Z_j are the components in phase with $\cos\omega_0 t$ and $\sin\omega_0 t$, respectively; R_{jK} and θ_{jK} are interpreted as the *envelope* and *phase* of the sum (2.6b) when the wave is truly narrow-band.

We have the following special models:

I. *Periodic, non-overlapping impulse noise:* The elementary pulses are assumed to occur at random within period intervals of length T_0. Thus, we have $t_{k(j)}'=k(j)T_0+\epsilon_{k(j)}$, where $\epsilon_{k(j)}$ is a random epoch representing the time interval between the beginning of a typical pulse and its period interval. Since there is no overlapping of impulses, we require that

$$e_j(t_{j'}-t_{k(j)}';r_{k(j)})=e_j(t_{j'};r_{k(j)}),$$
$$[k(j)T_0+\epsilon_{k(j)}<t_{j'}<$$
$$=0\text{ elsewhere, }(k(j)+1)T_0+\epsilon_{k(j)}+r_{k(j)}], \quad (2.9)$$

provided $0\leq r_k\leq T_0$. For no overlapping the epoch $\epsilon_{k(j)}$ is restricted by the relation $0\leq\epsilon_{k(j)}+r_{k(j)}\leq T_0$. If now we consider our interval $(0, T)$ containing exactly K periods $(K=N+1\geq 1)$, so that $T=(N+1)T_0$, $N\geq 0$, the basic probability $W_1(K)_T$ becomes $\delta_K{}^{N+1}$, since there are exactly $N+1$ events in $(0, T)$. The frequency functions for the random parameters $t_{k(j)}'$, $a_{k(j)}$, $\epsilon_{k(j)}$, $r_{k(j)}$, and $\psi_{k(j)}$ may be expressed as

$$w(\mathbf{a})=w(a_{0(j)}\cdots a_{K(j)};\ \epsilon_{0(j)}\cdots\epsilon_{k(j)};\cdots r_{k(0)}\cdots r_{K(j)};\cdots\psi_{k(j)}\cdots;) \cdot\prod_{k(j)=0}^{K}\delta(t_{k(j)}'-\epsilon_{k(j)}-k(j)T_0). \quad (2.10)$$

Substituting this expression in Eq. (2.4) and using the explicit relation for the basic probability $W_1(K)_T$ give us at once from Eqs. (2.1)–(2.3)

$$W_l(X_1,t_1;\cdots;X_s,t_s)_T=\prod_{j=1}^{s}\delta(X_j-0)$$

for $K=0$, or

$$=\int_{-\infty}^{\infty}\frac{d\xi_1}{2\pi}\cdots\int_{-\infty}^{\infty}\frac{d\xi_s}{2\pi} \times\exp[-i\sum_{j=1}^{s}\xi_j X_j]F_l(\xi_1,t_1;\cdots\xi_s,t_s)_{(K\geq 1)}^{N(\geq 0)}, \quad (2.11a)$$

where the characteristic function is

$$F_l(\xi_1,t_1;\cdots\xi_j,t_j;\cdots;\xi_s,t_s)_{(K\geq 1)}^{N(\geq 0)} =\int_{(\mathbf{a})}w(a_{0(j)}\cdots_{N(j)})da_{0(j)}\cdots d\psi_{(N+1)(j)} \times\exp\{i\sum_{j=1}^{s}\xi_j\sum_{k(j)=0}^{N+1}a_{k(j)}e_j(t_{j'}-k(j)T_0 -\epsilon_{k(j)};r_{k(j)})\cos\psi_{k(j)}\}. \quad (2.11b)$$

For narrow-band waves one has similar results, in the $2s$ random variables Y_j and Z_j, namely,

$$W_l(Y_1,Z_1,t_1;\cdots;Y_s,Z_s,t_s)_T =\prod_{j=1}^{s}\delta(Y_j-0)\delta(Z_j-0),$$

for $K=0$, or

$$=\int^{\infty}\int_{-\infty}\frac{d\xi_1 d\eta_1}{(2\pi)^2}\cdots\int^{\infty}\int_{-\infty}\frac{d\xi_s d\eta_s}{(2\pi)^2} \times\exp\{-i\sum_{j=1}^{s}\xi_j Y_j-i\sum_{j=1}^{s}\eta_j Z_j\} \cdot F_l(\xi_1,\eta_1,t_1;\cdots;\xi_s,\eta_s,t_s)_{(K\geq 1)}^{N(\geq 0)}, \quad (2.12)$$

and the exponent of the characteristic function (2.11b) is appropriately modified with the help of Eqs. (2.8a), (2.8b) in place of Eq. (2.6a).

When the X_j are purely periodic, the epochs $\epsilon_{k(j)}$ may be set equal to a random epoch ϵ, which is the same for all k and which corresponds to random observations of the wave trains by the observer. Here ϵ is a purely random variable representing completely random ob-

servations, so that $w_2(\epsilon)$ is equal to $1/T_0$ for $0 \leq \epsilon \leq T_0$ and is zero for other values. Thus, instead of a fixed observer (in time) examining a phenomenon changing randomly in time with respect to him, the familiar roles of observer and observed are reversed: the observer is "randomized" with respect to the phenomenon. For waves whose time-dependence is already governed by a stochastic process, the latter approach provides nothing new: our results are independent of ϵ, as expected, but for periodic waves this device does enable us to calculate W_1 directly. As a simple example, let us suppose that $X_1(t_1)=a\cos\omega_0 t_1$, where a may be randomly distributed, and we ask for $W_1(X_1, t_1)$, where the interval $(0, T)$ becomes infinite and we assume a stationary process. Then $w(\mathbf{a})=w_1(a)w_2(\epsilon)$, and from Eqs. (2.11a) and (2.11b) we obtain directly

$$W_1(X_1, t_1)=\int_{-\infty}^{\infty}(d\xi/2\pi)$$
$$\times\exp(-iX_1\xi)\int_{-\infty}^{\infty}w_1(a)da\int_0^{T_0}(d\epsilon/T_0)$$
$$\times\exp[ia\cos(\omega_0 t_1+\epsilon)]$$
$$=\int_{-\infty}^{\infty}(d\xi/2\pi)\langle J_0(a\xi)\rangle\exp[-i\xi X_1] \tag{2.13a}$$
$$\left.\begin{aligned}&=[1/\pi(a^2-X_1^2)^{\frac{1}{2}}]_{\text{stat av}}, \quad -a<X_1<a,\\ &=0, \quad |X_1|>|a|\end{aligned}\right\}. \tag{2.13b}$$

Some of the higher order probability densities are readily obtained in this instance.[24] In general, however, the explicit evaluation of Eq. (2.11) for the wave train is very difficult.

II. *Nonperiodic, non-overlapping impulse noise:* As in the first model, there is no overlapping of the elementary impulses which constitute the random wave. However, instead of a single pulse per period interval, one assumes now impulses of random duration such that the termination of one pulse provides the starting-point for the next. This means that like (2.9) one has

$$\left.\begin{aligned}e_j(t_{j'}-t_{k(j)}';\, r_{k(j)})&=e_j(t_{j'};\, r_{k(j)}),\\ &\qquad t_{k(j)}'<t_{j'}<t_{k(j)+1}'\\ &=0 \text{ elsewhere}\end{aligned}\right\}, \tag{2.14}$$

where now

$$r_{k(j)}=t_{k(j)+1}'-t_{k(j)}', \quad \text{or} \quad t_{k(j)}'=t_{0(j)}'+\sum_0^{k(j)-1}r_l. \tag{2.15}$$

Thus, the probability laws governing the distribution of the $t_{k(j)}'$ determine the statistical properties of the durations $r_{k(j)}$, so that if $\phi_T(t_k', t_{k+1}')$ is the joint probability density for the occurrence of two (consecutive) events (i.e., leading edges of pulses) in $(0, T)$ at times t_k, t_{k+1}, respectively, the frequency function $W_T(r_k)$ for the duration of the kth interval is formally

$$w_T(r_k)=\int_{-\infty}^{\infty}\phi_T(t_k', t_k'+r_k)dt_k'$$
$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\phi_T(t_0', \cdots t_k', t_k'+r_k;$$
$$t_{k+2}'\cdots t_K')dt_0'\cdots dt_K'. \tag{2.16}$$

When the times t_k' are all independent—the case of usual interest—$w_T(r_k)$ is easily found. We have then a special example of the distribution of the *range*[25] r_k, which for our present model becomes

$$w_T(r_k)=2\int_0^{T-r_k}\phi_T(t_k')\phi_T(t_k'-r_k)dt_k'$$
$$(0\leq r_k\leq T), \tag{2.17}$$

where $\phi_T(t_k)$ is the basic probability of occurrence of the "event" t_k'. In the general case, one has

$$w(\mathbf{a})=w(a_{0(j)}\cdots a_{K(j)};\, r_{0(j)}\cdots r_{K(j)};$$
$$\psi_{0(j)}\cdots\psi_{K(j)};\, t_{0(j)}'\cdots t_{K(j)}')$$
$$\cdot\prod_{k(j)=0}^{K}\delta(t_{k(j)}'-t_{k(j)-1}'-r_{k(j)-1}). \tag{2.18}$$

The probability density W_l may be written

$$W_l(X_1, t_1;\cdots;X_s, t_s)_T=W_1(0)_T\prod_{j=1}^{s}\delta(X_j-0),$$

for $K=0$, or

$$=\sum_{K=1}^{\infty}W_1(K)_T\int_{-\infty}^{\infty}\frac{d\xi_1}{2\pi}\cdots\int_{-\infty}^{\infty}\frac{d\xi_s}{2\pi}$$
$$\times\exp[-i\sum_j^s\xi_jX_j]\int_{(\mathbf{a})}w(\mathbf{a})\mathbf{da}$$
$$\times\exp[i\sum_j^s\xi_j\sum_{k(j)=0}^{K}a_{k(j)}e_j(t_{j'}-t_{k(j)}';$$
$$r_{k(j)})\cos\psi_{k(j)}], \tag{2.19}$$

in which the last set of integrations (over the random parameters) is the characteristic function corresponing to W_l. For narrow-band disturbances we have expressions of the type (2.12), for the $2s$ random variables Y_j, $Z_j(j=1, 2, \cdots s)$, given by (2.8a) and subject to Eqs. (2.14) and (2.15) above. In stationary cases, $T\to\infty$, and the explicit form of $w(\mathbf{a})$ simplifies somewhat with regard to the distributions of $r_{k(j)}$ (and $t_{k(j)}'$).

[24] For details, see D. Middleton, Technical Report No. 111, Cruft Laboratory, Harvard University, October 23, 1950, Eqs. (2.14a, b).

[25] L. H. C. Tippett, Biometrika **17**, 364 (1925); also P. G. Hoel, *Introduction to Mathematical Statistics*, (John Wiley and Sons, Inc., New York, 1947), pp. 161–164.

III. *Nonperiodic, overlapping impulsive noise: poisson noise:* Of all special models, the one of greatest general use is that of the so-called poisson noise. Here the random wave consists of the (linear) superposition of a large number of individual impulses each of which, however, is assumed to occur independently of the others in the interval of time $(0, T)$. The significant feature of this model is that, unlike (I) or (II), overlapping may occur. As before, the wave forms of these elementary pulses are assumed to be identical, and their amplitudes (a_k), durations (r_k), and phases (ψ_k) can vary in a random fashion; their times of occurrence (t_k') are independent, purely random quantities. The particular wave forms, of course, are governed in practical situations by the frequency response of the circuits or the selective properties of the medium in which the disturbance is evolved and through which it eventually passes.

As has been shown by Hurwitz and Kac,[26] the assumptions on $a_{k(j)}$, $r_{k(j)}$, $\psi_{k(j)}$, $t_{k(j)}'$, and e_j define a poisson distribution for the K events, from which follows our name for this type of noise wave. Here K is allowed to take only the integral values $(0, 1, 2, \cdots K \cdots \infty)$, each independently of any other value; thus, we obtained the specific basic probability $W_1(K)_T$ that, out of an infinite number of similarly prepared systems, each of duration T and in each one of which there may be $(0, 1, 2, \cdots K, \cdots \infty)$ impulses, the probability that an interval will contain exactly K events is given by

$$W_1(K)_T = (\langle N\rangle)^K \exp(-\langle N\rangle)/K!, \qquad (2.20)$$

where $\langle N\rangle$ is the average number of impulses arriving in $(0, T)$ for the ensemble[27] of strips $(0, T)$. Equation (2.20) may be written equivalently $W_1(K)_T = (nT)^K e^{-nT}/K!$, where n is the average number of events per unit time determined from the ensemble by

$$n = \lim_{m\to\infty} [(K_1 + K_2 + \cdots + K_m)/mT] = \langle N\rangle/T, \quad (2.21)$$

m being the number of similarly prepared strips. It is assumed that n exists; even in the limit $T\to\infty$, when $\langle N\rangle$ becomes indefinitely large, one still has n equal to a finite quantity, a condition always satisfied in physical problems because of power limitations.

The characteristic function for poisson noise can now be obtained from Eqs. (2.1), (2.3), and (2.20). We have

$$\begin{aligned} F_l(\xi_1, t_1; \cdots; \xi_s, t_s)_T &= \sum_{K=0}^{\infty} W_1(K)_T F_l(\xi_1, t_1; \cdots; \xi_s, t_s)_{T,K} \\ &= \sum_{K=0}^{\infty} \frac{(nT)^K}{K!} e^{-nT} \prod_{j=1}^{s} \int_{(\mathbf{a})} \sum_{k=0}^{K} \\ &\quad \times w(a_{k(j)}, r_{k(j)}, \psi_{k(j)}, t_{k(j)}')\mathbf{da} \\ &\quad \times \exp\{i\xi_j a_{k(j)} e_j(t_{j'} - t_{k(j)}'; r_{k(j)}) \cos\psi_{k(j)}\} \\ &= \sum_{K=0}^{\infty} \frac{(nT)^K}{K!} e^{-nT} \left\{ \int_0^{\infty} dr \int_{-\infty}^{\infty} da \int_{-\infty}^{\infty} d\psi \right. \\ &\quad \times \int_0^T w(a, r, \psi, \tau_0) d\tau_0 \\ &\quad \left. \times \exp[ia \cos\psi \sum_{j=1}^{s} \xi_j e_j(t_{j'} - \tau_0, r)] \right\}^K, \quad (2.22) \end{aligned}$$

since the random parameters a_k, r_k, ψ_k, and t_k' are independent (for different k), with identical distributions for all k. It is this independency condition which permits the factoring of the $4K$-fold multiple integrals into a simple Kth power of a fourfold integral, a procedure which is not in general permitted in noise models I and II above. Substituting Eq. (2.22) into Eq. (2.1) and summing over all values of K give at once

$$\begin{aligned} W_l(X_1, t_1; \cdots X_s, t_s)_T &= \int_{-\infty}^{\infty} \frac{d\xi_1}{2\pi} \cdots \int_{-\infty}^{\infty} \frac{d\xi_s}{2\pi} \\ &\quad \times \exp\left\{ -i \sum_{j=1}^{s} \xi_j X_j + \langle N\rangle \int_0^{\infty} dr \int_{-\infty}^{\infty} da \right. \\ &\quad \times \int_{-\infty}^{\infty} d\psi \int_0^T w(a, r, \psi, \tau_0) \\ &\quad \left. \times \{\exp[ia \cos\psi \sum_j^s \xi_j e_j(t_{j'} - \tau_0; r)] - 1\} d\tau_0 \right\}, \quad (2.23) \end{aligned}$$

which is the generalization of Rice's result[14] (2.6–9), *et seq.*

Some simplification of the probability density (2.23) is possible if we use the general properties of the individual pulses e_j. First, if we specify arbitrarily that outside the fundamental interval $(0, T)$ there is no disturbance, we may safely extend the limits of integration over τ_0 to $\pm\infty$. A small error is introduced by ignoring the relatively rare pulses which fall within a mean pulse length of the ends of the interval, but this effect is ignorable in all practical cases, where the interval $(0, T)$ is long compared with the average duration $\langle r\rangle$ of a typical impulse. In the limit $T\to\infty$ this error vanishes $0(1/T)$. Second, we require that the pulse shape e_j be such that integration over τ_0 is con-

[26] H. Hurwitz and M. Kac, Am. Math. Stat. **15**, 173 (1944).

[27] The average over the ensemble of strips follows directly from the concept of probability or probability density, which are themselves defined by the data contained in the (infinite) ensemble of similarly prepared systems. One has, in general, to distinguish between statistical averages, where the averaging is performed with respect to probability density as weighting function, and time averages, performed with respect to a weighting function $1/T$, the reciprocal of the interval of time over which the quantity is defined. In the former instance one considers the properties of the ensemble at a set of times $(t_1, t_2 \cdots t_s)$, within the strip-length T, while in the latter instance attention is focused on a particular strip (m) for the entire time $(0, T)$. The two averages are equivalent (to within a set of member functions of the ensemble of probability zero) if the random process is ergodic, as a consequence of ergodic theory. See A. Khintchine, *Statistical Mechanics* (Dover Publications, New York, 1949), Chapter III; see also reference 12.

vergent for all T. The fact that physical pulses always contain finite energy insures this convergence. It is convenient next to make the following substitutions:

$$x=\beta(t_1-\tau_0);\quad x_j=\beta(t_j-t_1),\quad (j=1, 2, \cdots s),\quad (2.24)$$

where $\beta(=\beta_1)$ is described in terms of the mean duration of the first set of pulses according to the relation $\beta=\langle\tau\rangle^{-1}$; the other sets of events $e_j(j=2, \cdots s)$ also have their characteristic "shape-factors" β_j, which appear implicitly in e_j. Thus, from Eq. (2.24) one writes $h_j=e_j((x+x_j)/\beta; \tau)$. The second term in the exponent of Eq. (2.23) becomes now

$$\frac{\langle N\rangle}{\beta}\int_0^\infty d\tau\int_{-\infty}^{\infty}da\int_{-\infty}^{\infty}d\psi\int_{-\infty}^{\infty}w\left(a,\tau,\psi,t_1-\frac{x}{\beta}\right)$$

$$\times[\exp(ia\cos\psi\sum_{j=1}^{s}\xi_j h_j(x+x_j,\tau))-1]dx. \quad (2.25)$$

For stationary random processes the probability density $w(a, \tau, \psi, t_1-x/\beta)$ is independent of the choice of time origin, with the consequence that setting $t_1=0$, one has equivalently $w(a, \tau, \psi, -x/\beta)$. Furthermore, w is symmetrical in time ($\sim x$), since there can be no distinction statistically between "forward" or "backward" in time if the process is stationary. Thus, we see that

$$w(a,\tau,\psi,t_1-x/\beta)=w(a,\tau,\psi,-x/\beta)$$
$$=w(a,\tau,\psi,x/\beta). \quad (2.26)$$

These modifications (2.24)–(2.26) permit us to write finally for the probability density in the stationary case

$$W_l(X_1;\cdots;X_s,t_s)$$
$$=\int_{-\infty}^{\infty}\frac{d\xi_1}{2\pi}\cdots\int_{-\infty}^{\infty}\frac{d\xi_s}{2\pi}F_l(\xi_1;\xi_2;t_2;\cdots;\xi_s,t_s)$$
$$\times\exp[-i\sum_{j}^{s}\xi_j X_j], \quad (2.27)$$

where the characteristic function is explicitly

$$F_l(\xi_1;\xi_2,t_2;\cdots;\zeta_s,t_s)$$
$$=\exp\Big\{(\langle N\rangle/\beta)\int_0^\infty d\tau\int_{-\infty}^{\infty}da\int_{-\infty}^{\infty}d\psi\int_{-\infty}^{\infty}$$
$$\times w(a,\tau,\psi;x/\beta)$$
$$\times(\exp[ia\cos\psi\sum_{j}^{s}\xi_j h_j(x+x_j,\tau)]-1)dx\Big\}. \quad (2.28)$$

Because the process is stationary, the interval length $(0, T)$ no longer enters explicitly. If now we assume that the epochs t_k' (and hence x above) are *uniformly* distributed in the interval $(0, T)$ and are independent of the durations, amplitudes, and phases of the various impulses, we may write $w(a, \tau, \psi, x/\beta)=w(a, \tau, \psi)/T$, so that

$$\lim_{T\to\infty}\frac{\langle N\rangle}{\beta}w\left(a,\tau,\psi,\frac{x}{\beta}\right)=w(a,\tau,\psi)\cdot\lim_{T\to\infty}\left(\frac{\langle N\rangle}{\beta T}\right)$$
$$=\gamma_{III}w(a,\tau,\psi), \quad (2.29)$$

where γ_{III} is a dimensionless parameter whose magnitude determines the character of the noise.

For narrow-band waves the decomposition into the slowly-varying components Y_j and Z_j follows at once from Eq. (2.8a), and the probability density (2.23) is correspondingly modified, to which in turn may be applied the simplifications when $T\to\infty$, similar to those for the broad-band case considered in detail above.

We expect the character of the distribution functions of X_j to depend on the amount of overlapping among individual impulses. For very heavy overlapping it is well known that the distribution of the X_j is multivariate-normal (see Sec. 5) and that, consequently, values of X_j of the order of the rms amplitude (of X_j) have a significant probability of recurring. On the other hand, when there is little overlapping, there will be appreciable gaps (in time) between successive pulses, and accordingly small or zero amplitudes are the most likely to occur, with only occasional bursts that are comparable with the rms amplitude of a single pulse. Moreover, in the former instance we anticipate that the precise form of the elementary, independent transients and their individual statistics are unimportant as far as the nature of the distribution is concerned, because there are so many pulses (in any short interval Δt) that their individuality is lost in the combined effect. (This follows directly from the Central Limit Theorem of probability.[10]) In the latter case, however, the shape and statistical properties of the individual impulses are critical in determining the form of the probability densities W_l, and it is this fact that makes an explicit evaluation of W_l so much more difficult in general. (For the trivial cases of ignorable overlapping one can clearly apply conventional methods to a single representative impulse.)

Since the type of poisson noise depends on the "density" of impulses in any given time interval, we see from Eq. (2.29) and the preceding argument that the characteristic parameter

$$\gamma_{III}=(\text{average number of impulses/sec})\cdot(\text{mean duration of a typical impulse})$$
$$=n\langle\tau\rangle=\frac{n}{\beta}=\lim_{T\to\infty}\frac{\langle N\rangle}{\beta T}$$

essentially determines the statistical character of the noise, *viz.*, whether or not it be impulsive static, nearly

TABLE I

Phenomenon: poisson noise	Magnitude of γ_{III}	Character of distributions
1. *Impulsive random noise:*	O(0–10)	Depends on individual pulse shape and pulse statistics. Strong dependence on magnitude of γ_{III}
(a) Static; ignition noise; solar interference	$O(10^{-1})$	
(b) Underwater sound; reflections from randomly oriented objects moving relative to observer	$O(10^{-1})$	
(c) Speech (random "burst" model)[a]	$O(10^{-1})$	
2. *Nearly normal random noise:*	$O(10)-O(10^4)$	Normal distribution with one or more correction terms. These are of order $\gamma_{III}^{-\frac{1}{2}}$ or γ_{III}^{-1}, depending on whether or not the third moments exists ($\neq 0$). Noticeable to weak dependence on magnitude of γ_{III}
(a) Heavy atmospheric static (ionosphere)		
(b) Precipitation noise		
(c) Clutter,[b] sea waves, etc.		
(d) Underwater sound (heavy overlapping of impulses for randomly moving obstacles)		
(e) Window (not densely sown)		
(f) Solar static (ionosphere); sunspot conditions		
3. *Normal random noise:*	$O(10^4-\infty)$	Normal distribution; ignorable correction terms (γ_{III} enters only as a scale factor for the probability densities, whose *form* now does not depend on γ_{III})
(a) Shot noise		
(b) Photomultiplier noise		
(c) Thermal noise		
(d) Clutter,[b] (scattering from water droplets)		
(e) Barkhausen noise		
(f) Window; electronic interference; inherent tube noise		

[a] W. Davenport, *A Study of Speech Probability Distributions*, Technical Report No. 148, Research Laboratory of Electronics, Massachusetts Institute of Technology, 1950.
[b] Reference 8.

normal random noise, or fluctuation noise. Table I lists a variety of physical situations to which the poisson model applies, the order of magnitude of γ_{III} associated with the phenomenon in question, and the general nature of the distribution densities $W_1, W_2, \cdots$, describing the random process. We classify the phenomena in three principal divisions; note also the acoustical as well as the electronic applications. These classifications are not rigid, since the order of magnitude of γ_{III} will vary somewhat with the physical situation. In the case of most impulsive noise γ_{III} is $O(10^{-1})$ or less; overlapping is not too heavy, so that the resultant noise waves X_j (Eqs. (2.6a) or (2.6b), exhibit a more or less discrete variation with time. Impulsive noise of denser type ($\gamma_{III}=O(1–10)$) occurs less frequently; analytically, such types are difficult to handle. Of the disturbances of an intermediate density ($\gamma_{III}=O(10–10^4)$), precipitation noise and some kinds of clutter are perhaps the most important; the mathematical theory here is closely allied with the more familiar treatment of fluctuation noise ($\gamma_{III}\to\infty$), since the distribution functions are asymptotically described by (linear) combinations of a normal frequency function and its derivatives (see Sec. 5) and consequently present no essentially new problems in their formal application to linear or nonlinear systems.[7,23] To the third class ($\gamma_{III}\to\infty$) belong the shot noise, thermal, and Barkhausen noise, etc., which possess well-known normal distributions.[7] Here we can get some idea of how well a normal distribution represents an actual case if we observe that for a typical mean tube current of 10 mils, the average number of elementary impulses per second (n) arriving at the plate is 10^{-2} coulombs/$e=3.10^7$ esu/$4.8\cdot 10^{-10}$ esu$=O(10^{17})$. On the other hand, the approximate duration of the impulses $\langle r\rangle$ are $O(10^{-8})$ so that $\gamma_{III}=n/\beta=O(10^9)$; and therefore, the principal correction terms in the distribution, which are $O(\gamma_{III}^{-\frac{1}{2}})$, are less than 10^{-2} percent of the leading terms. For more intense currents the normal law is even more closely approximated.

As one of the few examples where explicit results for any value of γ_{III} can be found, let us consider the case of a train of overlapping rectangular impulses, where the amplitudes are distributed according to the gaussian law $w_1(a)=\exp[-(a-\langle a\rangle)^2/2\sigma^2]/(2\pi\sigma^2)^{\frac{1}{2}}$, with $\sigma^2=\langle a^2\rangle-\langle a\rangle^2$ as the variance, and $w_2(\psi)=\delta(\psi-0)$, while the distribution of the durations $w_3(r)$ may have any meaningful value. This model is useful when we can safely

neglect the leading and trailing edges of the elementary pulses, i.e., when the duration alone is significant. The characteristic function for such a train of overlapping dc impulses is found for $s=1$ to be

$$F_1(\xi)=\exp\left[\gamma_{III}\int_{-\infty}^{\infty}w_1(a)(e^{ia}-1)da\right]$$
$$=\exp[-\gamma_{III}]\exp(\gamma_{III}\exp[i\langle a\rangle\xi-\sigma^2\xi^2/2])$$
$$=\exp(-\gamma_{III})\sum_{n=0}^{\infty}\frac{\gamma_{III}^n}{n!}\exp(in\langle a\rangle\xi-n\sigma^2\xi^2/2),$$
$$(\beta\langle r\rangle=1). \quad (2.30)$$

The corresponding probability density is obtained directly by inversion of Eq. (2.30):

$$W_1(X)=\exp(-\gamma_{III})\left[\delta(X-0)+\sum_{n=1}^{\infty}\frac{\gamma_{III}^n}{n!}\frac{\exp[-(X-n\langle a\rangle)^2]/2n\sigma^2}{(2\pi n\sigma^2)^{\frac{1}{2}}}\right]. \quad (2.31)$$

Thus, for small γ_{III} the major contributions to the probability density occur for zero amplitudes. In a similar way one finds that the second-order characteristic function for overlapping rectangular pulses is

$$F_2(\xi_1,\xi_2;t)$$
$$=\exp\left\{\gamma_{III}\int_{-\infty}^{\infty}w_1(a)da\int_{|t|}^{\infty}w_2(r)dr\int_0^{\beta(r-|t|)}\times[\exp ia(\xi_1+\xi_2)-1]dx\right\}, \quad (2.32)$$

since $h(x,r)=1$, $0<x<\beta r$, and $h(x+\beta t,r)=1$, $-\beta|t|<x<\beta r-\beta|t|$, while h vanishes outside these limits. If we represent the (normalized) correlation function of the elementary impulses by

$$\rho(t)=\beta\int_{|t|}^{\infty}(r-|t|)w_2(r)dr, \quad (2.33)$$

we can write finally that

$$F_2(\xi_1,\xi_2;t)=\exp(-\gamma_{III})\sum_{n=0}^{\infty}\frac{[\gamma_{III}\rho(t)]^n}{n!}\times\exp(in\langle a\rangle(\xi_1+\xi_2)-n\sigma^2(\xi_1+\xi_2)^2/2), \quad (2.34)$$

so that again by inversion it follows directly that

$$W_2(X_1,X_2;t)=\exp[-\gamma_{III}]\times\left\{\delta(X_1-0)\xi(X_2-0)+\sum_{n=1}^{\infty}\frac{[\gamma_{III}\rho(t)]^n}{n!}\times\delta(X_2-X_1)\frac{\exp[-(X_2-n\langle a\rangle)^2/2n\sigma^2]}{(2\pi n\sigma^2)^{\frac{1}{2}}}\right\}. \quad (2.35)$$

The relatively simple results (2.30)–(2.35) occur only because of our idealization of the pulse shape, which in turn allows us to decompose the characteristic functions into a sum of n-fold products of identical elements, as indicated in Eqs. (2.30) and (2.34). For pulses whose amplitude changes with the duration of the pulse, this procedure breaks down, and it is no longer possible to obtain explicit results for W_1, W_2, etc., for all powers of γ_{III}. However, one can still obtain the first few terms (up to $O(\gamma_{III}^3)$) in the general power series development, so that a satisfactory representation of low density impulsive noise is feasible for arbitrary pulse shapes. The general series expression for W_l is obtained directly from Eqs. (2.27) and (2.28). For a stationary process in which the basic pulses occur with a uniform distribution of arrival times (t_k), so that Eq. (2.29) applies, we write $\Phi_l(\xi_1;\xi_2,t_2;\cdots;\xi_s,t_s)$ for the coefficient of γ_{III} in the exponent of the characteristic function (2.28). Noting that

$$\Phi_l(0;0,t_2\cdots;0,t_s)=0, \quad (2.36)$$

we then have

$$W_l(X_1;X_2,t_2;\cdots;X_s,t_s)=\prod_{j=1}^{s}\delta(X_j-0)+\sum_{n=1}^{\infty}\frac{\gamma_{III}^n}{n!}\Lambda_n^{(l)}(X_1;X_2,t_2;\cdots;X_s,t_s), \quad (2.37)$$

where

$$\Lambda_n^{(l)}(X_1;X_2,t_2;\cdots X_s,t_s)=\int_{-\infty}^{\infty}\frac{d\xi_1}{2\pi}\cdots\int_{-\infty}^{\infty}\frac{d\xi_s}{2\pi}\Phi_l(\xi_1;\xi_2,t_2\cdots;\xi_s,t_s)^n\times\exp(-i\sum_{j=1}^{s}\xi_jX_j). \quad (2.38)$$

The nth term in the expansion of the characteristic function is simply $\gamma_{III}^n\Phi_l^n/n!$

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 9 SEPTEMBER, 1951

On the Theory of Random Noise. Phenomenological Models. II*

DAVID MIDDLETON
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received March 30, 1951)

Part II is devoted to the detailed evaluation of the first and second (second-order) moments, namely, average values and the cross- and autocorrelation functions for stationary random noise waves for models I–III, whose statistical description is given in Part I. From these in turn are determined the cross- and autospectral intensities governing the distribution of energy in the random disturbances. Finally, the asymptotically normal distributions characteristic of the high density poisson noise (III) are evaluated, together with higher order correction terms and their associated characteristic functions.

INTRODUCTION

IN the preceding two sections (Part I) we have constructed[1] the general probability densities $W_l(X_1, t_1 \cdots; X_s, t_s)$ for the three macroscopic noise models of greatest interest: (I) non-overlapping, periodic impulsive noise, (II) non-overlapping nonperiodic impulsive noise, and (III) poisson noise, or overlapping, nonperiodic pulses. The first two have many applications in communication theory,[2–5] while the last (III) covers a large class of electronic noise sources (see Table I, Part I). The following sections give us explicitly the quantities of principal physical interest: (a) the mean or average amplitude of the noise disturbance; (b) the mean intensity (proportional to the power) of the fluctuating noise waves; and (c) the spectral distribution of this power.

3. MOMENTS, CORRELATION FUNCTIONS, AND SPECTRA: GENERAL REMARKS

The moments of the distribution are of central importance: in linear systems the first and second moments of the random waves are needed to describe the above quantities of chief physical interest (see above and Sec. 1), while for nonlinear systems, moments of all orders may be required.[4,5] One can easily obtain the Qth moment $\mu_{q(1)\cdots q(s)}{}^{(Q)}$ directly from the characteristic function (2.4) by differentiation:[6]

$$\mu_{q(1)\cdots q(s)}{}^{(Q)} = \langle X_1{}^{q(1)} \cdots X_s{}^{q(s)} \rangle$$

$$= (-i)^Q \frac{\partial^Q}{\partial \xi_1{}^{q(1)} \cdots \partial \xi_s{}^{q(s)}} \times F_l(\xi_1, t_1; \cdots; \xi_s, t_s)_T \Big|_{\xi_1=\xi_2=\cdots\xi_s=0}$$

$$= \sum_{K=0}^{\infty} W_1(K)_T (-i)^Q \frac{\partial^Q}{\partial \xi_1{}^{q(1)} \cdots \partial \xi_1{}^{q(s)}} \cdot \int_{(\mathbf{a})} w(\mathbf{a}) d\mathbf{a} \exp[-\sum_{j=1}^{s} \xi_j X_j(t_{j'}; \mathbf{a})_K]. \quad (3.1)$$

The characteristic function itself has the Taylor series

$$F_l(\xi_1, t_1; \cdots; \xi_s, t_s)_T = 1 + \sum_{Q=1}^{\infty} \frac{(i\xi_1)^{q(1)} \cdots (i\xi_s)^{q(s)}}{q(1)! \cdots q(s)!} \mu_{q(1)\cdots q(s)}{}^{(Q)}. \quad (3.2)$$

In many cases useful auxiliary quantities are the *semi-invariants*[6] $\lambda_{m(1)\cdots m(s)}{}^{(M)}$ of the distribution, which are also found from the characteristic function by means of the defining relation

$$F_l(\xi_1, t_1; \cdots; \xi_s, t_s)_T = \exp\left\{ \sum_{M=1}^{\infty} \frac{(i\xi_1)^{m(1)} \cdots (i\xi_s)^{m(s)}}{m(1)! \cdots m(s)!} \lambda_{m(1)\cdots m(s)}{}^{(M)} \right\}. \quad (3.3)$$

In Eqs. (3.2) and (3.3) $\sum Q$ and $\sum M$ indicate the sum of all possible combinations of $Q = q(1) + \cdots q(s) \geq 1$, and $M = m(1) + \cdots + m(s) \geq 1$, respectively, for the particular values of Q or M chosen subject to the condition that $q(j) \geq 0$, $m(j) \geq 0$. For most of our present problems it is sufficient to confine our attention to stationary distributions of no higher than the second order ($l = 2$) (see Sec. 1); and as before, we have to distinguish between broad- and narrow-band disturbances.

* The research reported in this paper was made possible through support extended Cruft Laboratory, Harvard University, jointly by the Navy Department, Office of Naval Research, the Signal Corps of the U. S. Army, and the U. S. Air Force, under ONR Contract N5ori-76, T. O. 1.

[1] D. Middleton, J. Appl. Phys. **22**, 1143 (1951). (Additional references are included in this paper.)

[2] J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals* (McGraw-Hill Book Company, Inc., New York, 1950), M.I.T. Radiation Laboratory Series No. 24.

[3] James, Phillips, and Nichols, *Theory of Servo-Mechanisms* (McGraw-Hill Book Company, Inc., New York, 1947), M.I.T. Radiation Laboratory Series No. 25.

[4] Y. W. Lee, *Communication Applications of Correlation Analysis*, Symposium on Applications of Autocorrelation Analysis to Physical Problems, Woods Hole, June 13, 14 (1949), Department of the Navy (ONR). See also, D. Middleton, *Noise and Nonlinear Communication Problems*, *ibid.*

[5] D. Middleton, Quart. Appl. Math. **5**, 445 (1948); J. Appl. Phys. **19**, 817 (1948); Proc. Inst. Radio Engrs. **36**, 1467 (1948); J. Appl. Phys. **20**, 234 (1949); Quart. Appl. Math. **7**, 128 (1949); *ibid.*, **8**, 59 (1950).

[6] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, New Jersey, 1946).

(i) Broad-Band Waves

Here we let X_j $(j=1, 2, \cdots n)$ represent the j random variables formed from the j sets of elementary pulses e_j in the manner of Sec. 2, Eqs. (2.6a, b), such that $X_{j+n}=X_j$ at a time t later, there being $s=2n$ random variables altogether. The moments $\mu_{q(1)\cdots q(2n)}{}^{(Q)}$ and the semi-invariants $\lambda_{m(1)\cdots m(2n)}{}^{(M)}$ are related through Eqs. (3.2) and (3.3).

The first- and second-order terms $(M, Q=1, 2)$ are of chief interest; these may be conveniently expressed in matrix form. For the mean values ν_k $(k=1, \cdots 2n)$ and first-order semi-invariants $\lambda_{0\cdots1\cdots0}{}^{(1)}$ we accordingly write

$$\mathbf{v}=\|\nu_k\|=\begin{vmatrix}\nu_1\\ \nu_2\\ \cdot\\ \cdot\\ \cdot\\ \nu_{2n}\end{vmatrix}=\begin{vmatrix}\langle X_1\rangle\\ \cdots\\ \cdot\\ \cdot\\ \cdot\\ \langle X_{2n}\rangle\end{vmatrix}=\|\mu_k{}^{(1)}\|,$$

and

$$\mathfrak{L}^{(1)}=\begin{vmatrix}\lambda_{10\cdots0}\\ \lambda_{01\cdots0}\\ \cdot\\ \cdot\\ \cdot\\ \lambda_{0\cdots01}\end{vmatrix}, \quad (M, Q=1). \tag{3.4}$$

Because the random process to which X_j $(j=1, \cdots 2n)$ belongs is assumed to be stationary, the averages are independent of the time; one has, therefore, $\langle X_j\rangle = \langle X_{n+j}\rangle$, $(j=1, 2\cdots n)$. The column vectors $\mathbf{v}$ and $\mathfrak{L}^{(1)}$ may now be written compactly as

$$\mathbf{v}=\begin{vmatrix}\mathbf{v}_0\\ \mathbf{v}_0\end{vmatrix}, \quad \text{and} \quad \mathfrak{L}^{(1)}=\begin{vmatrix}\mathfrak{L}_0{}^{(1)}\\ \mathfrak{L}_0{}^{(1)}\end{vmatrix}, \tag{3.5}$$

where $\mathbf{v}_0$ and $\mathfrak{L}_0{}^{(1)}$ are column vectors with the elements $\|\nu_0\|=\|\langle X_j\rangle\|$, $\|\lambda_0\|=\|\lambda_{0\cdots1\cdots0}\|$. In a similar way the second moments M_{kl} $(k, l=1, 2\cdots2n)$ and semi-invariants $\lambda_{010\cdots01\cdots0}{}^{(2)}$ may be expressed

$$\mathbf{M}=\|M_{kl}\|=\|\langle X_kX_l\rangle\|=\|\mu_{kl}{}^{(2)}\|;$$

$$\mathfrak{L}^{(2)}=\begin{Vmatrix}\lambda_{200\cdots0} & \lambda_{110\cdots0} & \lambda_{101\cdots0}\cdots\lambda_{100\cdots01}\\ \lambda_{110\cdots0} & \lambda_{020\cdots0} & \lambda_{011\cdots0} \quad \cdots\\ & & \quad \cdot\\ & & \quad \cdot\\ & & \quad \cdot\\ \lambda_{101\cdots0} & \lambda_{011\cdots0} & \lambda_{002\cdots0}\\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \lambda_{10\cdots01} & & \lambda_{0\cdots02}\end{Vmatrix}. \tag{3.6}$$

Both $\mathbf{M}$ and $\mathfrak{L}^{(2)}$ are square $(2n\times2n)$ matrices. Comparison of coefficients of $(i\xi_1)^{q(1)}\cdots(i\xi_{2n})^{q(2n)}$ in Eqs. (3.2) and (3.3) when $M, Q=1, 2$ shows at once that the matrices of the averages $\langle X_j\rangle$ and the second moments $\langle X_kX_l\rangle$ $(j=k, l, =1, \cdots2n)$ are related to the matrices of the semi-invariants $\mathfrak{L}^{(1)}$, $\mathfrak{L}^{(2)}$ by

$$\mathbf{v}=\mathfrak{L}^{(1)}; \quad \mathbf{M}=\mathfrak{L}^{(2)}+\|\nu_k\nu_l\|=\|\mu_{kl}{}^{(2)}\|. \tag{3.7}$$

In many cases it is convenient to introduce the variance matrix $\boldsymbol{\sigma}$:

$$\boldsymbol{\sigma}=\mathbf{M}-\|\nu_k\nu_l\|=\mathfrak{L}^{(2)};$$

$$\boldsymbol{\sigma}=\|\sigma_{kl}\|=\|\langle(X_k-\langle X_k\rangle)(X_l-\langle X_l\rangle)\rangle\|, \quad (k, l=1, 2\cdots2n). \tag{3.8}$$

Not only are the square matrices symmetrical, but they can be decomposed into a symmetrical (2×2) matrix, each of whose elements is an $(n\times n)$ matrix (which, however, is not necessarily symmetrical). This decomposition is possible because we are dealing here with a stationary, second-order distribution, where only the time difference $t(=t_2-t_1)$ between initial and final states enters. Accordingly, Eq. (3.6) may be written

$$\mathbf{M}=\begin{Vmatrix}\mathbf{R}(0) & \mathbf{R}(t)\\ \mathbf{R}(t) & \mathbf{R}(0)\end{Vmatrix}, \quad \text{and} \quad \mathfrak{L}^{(2)}=\begin{Vmatrix}\boldsymbol{R}(0) & \boldsymbol{R}(t)\\ \boldsymbol{R}(t) & \boldsymbol{R}(0)\end{Vmatrix}, \tag{3.9}$$

$$\text{with } \boldsymbol{R}(\mathbf{t})=\|\mathfrak{R}_{kl}(t)\|$$

in which

$$\mathbf{R}(\mathbf{t})=\|R_{kl}(t)\|=\|\langle X_k(t_1)X_l(t_1+t)\rangle\|=\mathbf{R}'(-\mathbf{t}), \quad (k, l=1, 2\cdots n), \tag{3.9a}$$

where the prime indicates the transposed matrix. From Eqs. (3.6) and (3.8) one can write for the k, lth element

$$R_{kl}(t)=\mathfrak{R}_{kl}(t)+\nu_k\nu_l=\sigma_{kl}(t)+\nu_k\nu_l. \tag{3.10}$$

The quantities $R_{kl}(t)$ are correlation functions; when $k=l$, one has the set of *autocorrelation functions* $R_{kk}(t)$, $(k=1\cdots n)$ for the random variables $X_1\cdots X_n$. On the other hand, when $k\neq l$, one has instead the *cross-correlation functions* $R_{kl}(t)$ for these variables. From Eqs. (3.9a) and (3.10) we observe that $R_{kl}(t)=R_{lk}(-t)$, and the autocorrelation functions are therefore always even functions of t, as expected for stationary processes.

(ii) Narrow-Band Waves

We have now to consider the distribution of the in- and out-of-phase components Y_j, Z_j; see Eqs. (2.7) and (2.8). Here one has a $4n$-dimensional probability density in Y_j and Z_j $(j=1, \cdots2n)$, since a narrow-band disturbance may alternatively be described by a pair of random variables. As before, Y_j and Z_j $(j=1\cdots n)$ correspond to the variables at some initial time and Y_j, Z_j $(j=n+1, \cdots, 2n)$ to these variables respectively, at a later time t. The mean values and first semi-invariants may be expressed as column matrices:

$$\mathbf{v}=\begin{vmatrix}\mathbf{v}^{(y)}\\ \mathbf{v}^{(z)}\end{vmatrix}=\mathfrak{L}^{(1)}=\begin{vmatrix}\mathfrak{L}_y{}^{(1)}\\ \mathfrak{L}_z{}^{(1)}\end{vmatrix}=\begin{vmatrix}\|\langle Y_k\rangle\|\\ \|\langle Z_k\rangle\|\end{vmatrix}, \tag{3.11}$$

while the correlation matrix $\mathbf{M}$, Eq. (3.9), becomes

$$\mathbf{M}=\begin{Vmatrix}\mathbf{R}(0) & \mathbf{R}(t)\\ \mathbf{R}'(t) & \mathbf{R}(0)\end{Vmatrix},\ \text{with}\ \mathbf{R}(t)=\begin{Vmatrix}\mathbf{R}_{yy}(t) & \mathbf{R}_{yz}(t)\\ \mathbf{R}_{zy}(t) & \mathbf{R}_{zz}(t)\end{Vmatrix}. \quad (3.12)$$

Note that now $\mathbf{M}$ is not symmetrical in $\mathbf{R}(t)$, unlike the simpler case (3.9), since $\mathbf{R}(t)$ is itself not in general symmetrical, because in $\mathbf{R}_{yz}(t)$ the averages $\langle Y_k Z_l\rangle$ and $\langle Z_k Y_l\rangle$ are not equal and do not vanish. The elements $\mathbf{R}_{yy}(t)$, $\mathbf{R}_{yz}(t)$, etc., of the matrix $\mathbf{R}(t)$ are themselves square $(n\times n)$ matrices such that

$$\mathbf{R}_{yy}(t)=\|R_{kl}^{(yy)}(t)\|=\|\langle Y_k(t_1)Y_l(t_1+t)\rangle\|,$$

$$\mathbf{R}_{yz}(t)=\|R_{kl}^{(yz)}(t)\|=\|\langle Y_k(t_1)Z_l(t_1+t)\rangle\|$$

$$=\mathbf{R}'_{zy}(-t), \quad (3.13)$$

$$\mathbf{R}_{zz}(t)=\|R_{kl}^{(zz)}(t)\|=\|\langle Z_k(t_1)Z_l(t_1+t)\rangle\|,$$

$$(k, l, \cdots n).$$

Similarly, one finds the matrix of the semi-invariants $\mathfrak{a}^{(2)}$ to be

$$\mathfrak{a}^{(2)}=\begin{Vmatrix}\boldsymbol{R}(0) & \boldsymbol{R}(t)\\ \boldsymbol{R}(t) & \boldsymbol{R}(0)\end{Vmatrix},\ \text{with}\ \boldsymbol{R}(t)=\begin{Vmatrix}\boldsymbol{R}_{yy}(t) & \boldsymbol{R}_{yz}(t)\\ \boldsymbol{R}_{zy}(t) & \boldsymbol{R}_{zz}(t)\end{Vmatrix}. \quad (3.14)$$

Since Eq. (3.8) applies here also, we may write

$$R_{kl}^{(yy)}(t)=\mathfrak{R}_{kl}^{(yy)}(t)+\nu_k^{(y)}\nu_l^{(y)}$$

$$=\sigma_{kl}^{(yy)}(t)+\nu_k^{(y)}\nu_l^{(y)}, \quad (k, l=1, \cdots n) \quad (3.15)$$

with corresponding expressions for $R_{kl}^{(yz)}(t)$ and $R_{kl}^{(zz)}(t)$.

(iii) Spectra

The theorem of Wiener[7] and Khintchine[8] is used to obtain the various spectral energy distributions associated with the correlation functions $R_{kl}(t)$: the mean (auto) spectral density $W_{kk}(f)$ of a time-dependent disturbance is the cosine fourier transform of the corresponding autocorrelation function $R_{kk}(t)$. For the set of $k(k=1, \cdots, n)$ random variables one has therefore the pairs of relations

$$W_{kk}(f)=4\int_0^\infty R_{kk}(t)\cos\omega t dt, \quad (\omega=2\pi f), \quad (3.16a)$$

and

$$R_{kk}(t)=\int_0^\infty W_{kk}(f)\cos\omega t df. \quad (3.16b)$$

In a similar fashion the Wiener-Khintchine theorem is extended to include the cross-spectral intensities $W_{kl}(f)$. The analog of (3.16) in this instance is

$$W_{kl}(f)=2\int_{-\infty}^\infty R_{kl}(t)e^{i\omega t}dt,$$

and

$$R_{kl}(t)=\frac{1}{2}\int_{-\infty}^\infty W_{kl}(f)e^{-i\omega t}df, \quad (3.17)$$

where $R_{kl}(t)$ is found from Eq. (3.9a) or Eq. (3.10). For narrow-band waves one has alternatively from Eqs. (2.8) and (3.10) the equivalent relations

$$\begin{aligned}R_{kl}(t)&=\langle X_k(t_1)X_l(t_1+t)\rangle\\ &=[\{Y_k(t_1)\cos\omega_0 t_1+Z_k(t_1)\sin\omega_0 t_1\}\\ &\quad\times\{Y_l(t_1+t)\cos\omega_0(t_1+t)\\ &\quad+Z_l(t_1+t)\sin\omega_0(t_1+t)\}]_{\text{stat av}}\\ &=\tfrac{1}{2}\{R_{kl}^{(yy)}(t)+R_{kl}^{(zz)}(t)\}\cos\omega_0 t\\ &\quad+\tfrac{1}{2}\{R_{kl}^{(yz)}(t)-R_{kl}^{(zy)}(t)\}\sin\omega_0 t, \end{aligned} \quad (3.18)$$

from which the spectra follow according to (3.17).† Note that $R_{kk}(0)$ gives the mean total power in the disturbance, while $R_{kk}(\infty)$ yields the intensity in the dc or periodic terms. Consequently, the power associated with the continuous portion of the spectrum is $R_{kk}(\infty)-R_{kk}(0)$.

4. MOMENTS, CORRELATION FUNCTIONS, AND SPECTRA: SPECIFIC NOISE MODELS

(i) Broad-Band Waves

We give explicitly the first and second moments for the second-order distribution of the $2n$ stationary random variables X_j $(j=1, \cdots 2n)$ in the instances of the three principal noise models considered in the preceding sections. For the first case we have

(I) *Non-overlapping, periodic impulse noise*: Application of Eqs. (3.4) and (3.1) to the characteristic function (2.11b) gives at once (for the initial time $t_{j'}=t_0$)

$$\mu_k^{(1)}=\nu_k(t_0)=\sum_{m=0}^N\langle a_m\cos\psi_m e_k(t_0-mT_0-\epsilon_m; r_m)\rangle$$

$$=\sum_{m=0}^N\left\{a_m\cos\psi_m\int_{-\infty}^\infty S_k(f, r_m)\right.$$

$$\left.\times\exp[i\omega t_0-i\omega(mT_0+\epsilon_m)]df\right\}_{\text{stat av}}, \quad (4.1)$$

where

$$S_k(f, r_m)=\int_{-\infty}^\infty\exp(-i\omega t')e_k(t', r_m)dt' \quad (4.2)$$

is the amplitude spectral intensity of a typical elementary impulse (for the kth member of the set X_j, $(j=1, \cdots k\cdots 2n)$). We now let the interval $(0, T)$ become infinitely long, so that $(N+1)T_0=T\to\infty$. Since we are dealing with a stationary process, time and ensemble averages may be safely interchanged for all practical purposes (see note, Sec. 2, III), and we can

[7] N. Wiener, Acta Math. **55**, 117 (1930).
[8] A. Khintchine, Math. Ann. **109**, 604 (1934).

† The averages in Eq. (3.18) over the random amplitudes Y_k, Z_l, etc., and over the trigonometric terms $\cos\omega_0(t_1+t)$, etc., are performed separately, since the random variables are statistically independent of the periodic components.

now write Eq. (4.1) as

$$\nu_k = \lim_{N\to\infty} \sum_{m=0}^{N} \frac{a_m \cos\psi_m}{T_0(N+1)} \int_{-\infty}^{\infty} dt_0 \int_{-\infty}^{\infty} df S_k(f, r_m) \times \exp[i\omega t_0 - i\omega(mT_0+\epsilon_m)]. \quad (4.3)$$

Again because of the stationarity of the process, the range of t_0 may be safely extended to $-\infty$ as well as to $+\infty$, when $(0, T)$ becomes indefinitely long, X_k vanishing at all times outside this (infinite) interval. Noting that

$$\int_{-\infty}^{\infty} \exp(i\omega t_0) dt_0$$

is the singular function $\delta(f-0)$, we have, finally,

$$\nu_k = \frac{1}{T_0} \lim_{N\to\infty} \sum_{m=0}^{N} \frac{a_m \cos\psi_m}{N+1} S_k(0, r_m)$$

$$= \frac{1}{T_0} \langle a \cos\psi S_k(0, r)\rangle \quad (4.4a)$$

$$= \frac{1}{T_0} \int_{(\mathbf{a})} w_1(a, r, \psi) a \cos\psi S_k(0, r) \mathbf{da}, \quad (4.4b)$$

on replacing the time average by the equivalent ensemble average; w_1 is here the first-order probability density for the random parameters. As expected, the mean value ν_k is now independent of the (initial) times t_0, and of the random "jitter" ϵ. We remark that the "density" parameter γ, introduced for the poisson noise described in Sec. 2,[1] may be similarly defined here. One has, specifically, $\gamma_I = [(N+1)/T]\langle r\rangle = \langle r\rangle/T_0 \leq 1$. For rectangular impulses, then, ν_k is equal to $\gamma_I \langle a \cos\psi\rangle$.

The second moments are of chief interest. Again differentiating the characteristic function according to Eq. (3.1), we find that here for the initial and final times $t_{j'} = t_0,\ t_0+t$:

$$M_{kl} = \mu_{kl}^{(2)}$$

$$= \sum_{m=0}^{N} \sum_{n=0}^{N} \langle a_m a_n \cos\psi_m \cos\psi_n \times e_k(t_0 - mT_0 - \epsilon_m; r_m) e(t_0+t-nT_0-\epsilon_n; r_n)\rangle$$

$$= \sum_{m=0}^{N} \sum_{n=0}^{N} \Bigg[a_m a_n \cos\psi_m \cos\psi_n \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} df df' S_k(f, r_m) S_l(f', r_n) \times \exp\{i(\omega+\omega')t_0 + i\omega' t - i\omega(mT_0+\epsilon_m) - i\omega'(nT_0+\epsilon_n)\} \Bigg]_{\text{stat av.}} \quad (4.5)$$

For stationary systems, on letting $T\to\infty$ again, we may interchange time and ensemble averages. When this has been done, we see that now the time average introduces the integral

$$\int_{-\infty}^{\infty} \exp i(\omega'+\omega)t_0 dt_0 = \delta(f'+f),$$

so that by reasoning similar to that above for ν_k we have the cross-correlation function

$(= M_{kl})$:

$$R_{kl}(t) = \lim_{N\to\infty} \frac{1}{(N+1)T_0} \sum_{m=0}^{N} \sum_{n=0}^{N} a_m a_n \cos\psi_m \cos\psi_n \times \int_{-\infty}^{\infty} df S_k(f, r_m) S_l(f, r_m)^* \times \exp[-i\omega t - i\omega(m-n)T_0 - i\omega(\epsilon_m - \epsilon_n)]. \quad (4.6)$$

For the real waves examined here,

$$S_{k,l}(-f, r) = S_{k,l}(f, r)^*.$$

We write now

$$A_m^{(k)}(\omega, \mathbf{a}) = a_m \cos\psi_m S_k(f, r_m) \exp(-i\omega\epsilon_m);$$
$$A_n^{(l)}(\omega, \mathbf{a}) = a_n \cos\psi_n S_l(f, r_n) \exp(-i\omega\epsilon_n). \quad (4.7)$$

Letting $j = n-m$, we can rearrange the sums in Eq. (4.6) to obtain finally the desired general result

$$R_{kl}(t) = \frac{1}{T_0} \sum_{j=-\infty}^{\infty} \int_{-\infty}^{\infty} R_{kl}(j; \omega)_d \times \exp[-i\omega(t-jT_0)] df = R_{lk}(-t), \quad (4.8)$$

where

$$R_{kl}(j; \omega)_d \equiv \lim_{N\to\infty} \frac{1}{N+1} \sum_{m=0}^{N} A_m^{(k)}(\omega; \mathbf{a}) A_{m+j}^{(l)}(\omega; \mathbf{a})^*$$

$$= R_{lk}(-j; \omega)_d^* = R_{lk}(j; -\omega)_d^* \quad (4.9a)$$

$$= \langle [A_1^{(k)}(\omega; \mathbf{a}_1) A_2^{(l)}(\omega; \mathbf{a}_2)^*]_j \rangle$$

$$= \int_{\mathbf{a}_1, \mathbf{a}_2} \mathbf{da}_1 \mathbf{da}_2 A_1^{(k)}(\omega, \mathbf{a}_1) \times A_2^{(l)}(\omega, \mathbf{a}_2)^* w_2(\mathbf{a}_1, \mathbf{a}_2; jT_0), \quad (4.9b)$$

this last because of stationarity. The subscripts 1 and 2 in Eq. (4.9b) refer to the initial (m) and final $(m+j)$ states, separated by the time interval $t_2 - t_1 = jT_0$; note that, unlike $R_{kl}(t)$, which is a real quantity, $R_{kl}(j, \omega)_d$ is in general complex. We also observe from Eq. (4.2) that

$$\int_{-\infty}^{\infty} S_k(f, r_m) S_l(f, r_{m+j})^* \times \exp[-i\omega(t - jT_0 - \epsilon_{m+j} + \epsilon_m)] df$$

$$= \int_{-\infty}^{\infty} e_k(t_0; r_m) e(t_0+t-jT_0-\epsilon_{m+j}+\epsilon_m; r_{m+j}) dt_0$$

$$= 2\rho_{kl}(t - jT_0 - \epsilon_{m+j} + \epsilon_m; r_m, r_{m+j}), \quad (4.10)$$

in which ρ_{kl} is the (normalized) cross-correlation function for the elementary impulses e_k, e_l, so that we can express Eq. (4.8) in the alternative form

$$R_{kl}(t)=\frac{2}{T_0}\sum_{j=-\infty}^{\infty}\langle a_1 a_2\cos\psi_1\cos\psi_2 \times\rho_{kl}(t-jT_0-\epsilon_2+\epsilon_1;r_1,r_2)\rangle. \quad (4.11)$$

The mean cross-spectral density $W_{kl}(f)$ is obtained at once from Eq. (4.8) with the aid of Eq. (3.17). We have

$$W_{kl}(f)=\frac{2}{T_0}\sum_{j=-\infty}^{\infty}R_{kl}(j;\omega)_d\exp(ij\omega T_0)$$
$$=\frac{2}{T_0}\sum_{j=-\infty}^{\infty}\langle[A_1^{(k)}(\omega,\mathbf{a}_1)A_2^{(l)}(\omega,\mathbf{a}_2)^*]_j\rangle \cdot \times\exp(ij\omega T_0). \quad (4.12)$$

A special case of considerable usefulness arises when the parameters $\mathbf{a}$ belong to a purely random process; then $\mathbf{a}_1$ and $\mathbf{a}_2$ are independent, and do not involve the time difference $t_2-t_1=jT_0$. In such instances the averages over $A_1^{(k)}A_2^{(l)*}$ factor, and we are able to write finally for Eq. (4.8) the mean intensity spectrum, with the help of Eq. (3.17),

$$W_{kl}(f)=\frac{2}{T_0}\{\langle A_1^{(k)}(\omega,\mathbf{a})A_1^{(l)}(\omega,\mathbf{a})^*\rangle -\langle A_1^{(k)}(\omega,\mathbf{a})\rangle\langle A_1^{(l)}(\omega,\mathbf{a})^*\rangle + \sum_{j=-\infty}^{\infty}\exp(ij\omega T_0) \times[\langle A_1^{(k)}(\omega,\mathbf{a})\rangle\langle A_1^{(l)}(\omega,\mathbf{a})^*\rangle]_j\}. \quad (4.13)$$

As an interesting example, let us consider the case where the elementary pulses are of fixed duration (r_0), and where a, r, ψ, ϵ are independent, so that $w_1(a,r,\psi,\epsilon)=w_1(a)w_2(\epsilon)\delta(r-r_0)w_3(\psi)$. Writing

$$b(f)=\int_{-\infty}^{\infty}\exp(-i\omega\epsilon)w_2(\epsilon)d\epsilon,$$

we have from Eq. (4.13),

$$W_{kl}(f)=(2/T_0)S_k(f,r_0)S_l(f,r_0)^* \times\left\{\langle a^2\rangle\langle\cos^2\psi\rangle-\langle a\rangle^2\langle\cos\psi\rangle^2|b(f)|^2 +\langle a\rangle^2\langle\cos\psi\rangle^2\frac{|b(f)|^2}{T_0}\sum_{m=0}^{\infty}\delta\left(f-\frac{m}{T_0}\right)\right\}, \quad (4.14)$$

where the series $\sum_{j=-\infty,+\infty}\exp(ij\omega T_0)$ in Eq. (4.13) is equal to $\sum_{m=0,\infty}\delta(f-m/T_0)/T_0$, as shown in the Appendix of an earlier paper.[9] Our result (4.14) is one generalization of two expressions given by Lawson and Uhlenbeck[2] (pp. 43, 44). Note that one has besides the continuous spectrum a series of periodic components located at $f=mf_0=m/T_0$, the intensities of which are influenced not only by the elementary pulse shapes, but through $b(f)$ by the distribution of the "jitter" or displacement ϵ.

Our expressions (4.8), (4.11), (4.12) may also be used when the phases are cumulative, i.e., when the phase of the impulse in a given period interval is the sum of the previous phases and of the phase change characteristic of this period, namely, $\psi_m=\sum_{\lambda=0,m}\phi_\lambda$. The simplest, and usually most interesting cases, occur when ϕ_λ, r, a, and ϵ are all independent from period to period, although r, a, and ϵ may be functionally related.‡ Then one has for the typical jth (≥ 1) term in the summation in (4.13) on replacing ensemble averages by time averages

$$\lim_{N\to\infty}\frac{1}{N+1}\sum_{m=0}^{N}C_m^{(k)}(\omega,\mathbf{a})C_{m+j}^{(l)}(\omega,\mathbf{a})^* \times\tfrac{1}{4}\{\exp(2i\sum_{\lambda=0}^{m}\phi_\lambda+i\sum_{m+1}^{m+j}\phi_\lambda) +\exp(-2i\sum_{0}^{m}\phi_\lambda-i\sum_{m+1}^{m+j}\phi_\lambda) +\exp(i\sum_{m+1}^{m+j}\phi_\lambda)+\exp(-i\sum_{m+1}^{m+j}\phi_\lambda)\}, \quad (4.15)$$

where now $C_m^{(k)}$, $C_{m+j}^{(l)}$ are found from Eq. (4.7), according to the following:

$$A_m^{(k)}(\omega,\mathbf{a})=C_m^{(k)}(\omega,\mathbf{a})\cos\psi_m;$$
$$A_{m+j}^{(l)}(\omega,\mathbf{a})=C_{m+j}^{(l)}(\omega,\mathbf{a})\cos\psi_{m+j}. \quad (4.15a)$$

Since a, r, ϕ, ϵ are assumed independent, we find after replacing the time by ensemble averages in Eq. (4.15) for $(j\geq 1)$

$$\tfrac{1}{4}\langle C_1^{(k)}(\omega,\mathbf{a})\rangle\langle C_1^{(l)}(\omega,\mathbf{a})^*\rangle \times\sum_{j=1}^{\infty}\exp(ij\omega T_0)[(\langle\eta^2\rangle+1)(\langle\eta\rangle)^j +(\langle\eta^{*2}\rangle+1)(\langle\eta^*\rangle)^j] \quad (4.16)$$

in which $\eta=e^{i\phi}$, so that

$$\langle\eta\rangle=\int_0^{2\pi}e^{i\phi}w_2(\phi)d\phi,$$

and

$$\langle\eta^2\rangle=\int_0^{2\pi}e^{2i\phi}w_2(\phi)d\phi, \quad (\langle\eta^k\rangle\leq 1). \quad (4.17)$$

[9] D. Middleton, Cruft Laboratory Technical Report No. 111, October 23, 1951.

‡ Here, although we assume r, a, and ϵ independent, in order to satisfy the condition $r+\epsilon\leq T_0$ for no overlapping, the allowed values of r and ϵ must fall in the triangle $(r,\epsilon)=(0,0)$, $(0,T_0)$, $(T_0,0)$.

Excluding for the moment those values of $\langle\eta\rangle \exp(i\omega T_0)$ which have the value unity (as do $\langle\eta^*\rangle \exp(i\omega T_0)$, $\langle\eta^*\rangle \exp(-i\omega T_0)$, and $\langle\eta\rangle \exp(-i\omega T_0)$ simultaneously), we may sum the geometric series in Eq. (4.16) at once to get

$$\exp(i\omega T_0)\{\langle\eta\rangle[(\langle\eta^2\rangle+1)/(1-\langle\eta\rangle \exp(i\omega T_0))] + \langle\eta^*\rangle(\langle\eta^{2*}\rangle+1)/[1-\langle\eta^*\rangle \exp(i\omega T_0))]\}.$$

Repeating this procedure for $j(\leq -1)$ gives Eq. (4.16), except that now l and k are interchanged and $\exp(i\omega T_0)$ is replaced by $\exp(-i\omega T_0)$. For $j=0$ one has directly the joint average

$$(\langle\eta^2\rangle+\langle\eta^{2*}\rangle+2)\langle C_1^{(k)}(\omega, \mathbf{a})C_1^{(l)}(\omega, \mathbf{a})^*\rangle/4.$$

Combining these results according to Eq. (4.8) yields, finally, the cross-correlation function for the continuous part of the corresponding spectral distribution. We have

$$R_{kl}(t)=\frac{1}{4T_0}\int_{-\infty}^{\infty}\exp(-i\omega t)df \times\Big\{(\langle\eta^2\rangle+\langle\eta^2\rangle^*+2)\langle C_1^{(k)}(\omega, \mathbf{a})C_1^{(l)}(\omega, \mathbf{a})^*\rangle +2\mathcal{R}e\langle C_1^{(k)}(\omega, \mathbf{a})\rangle\langle C_1^{(l)}(\omega, _t\mathbf{a})^*\rangle \times\exp(i\omega T_0)\Big[\frac{\langle\eta\rangle(\langle\eta^2\rangle+1)}{1-\langle\eta\rangle \exp(i\omega T_0)} +\frac{\langle\eta^*\rangle(\langle\eta^{2*}\rangle+1)}{1-\langle\eta\rangle^* \exp(i\omega T_0)}\Big]\Big\}. \quad (4.18)$$

There is in general *no* constant or periodic contribution to the correlation function, and hence no steady or periodic component in the spectrum. A simple example of some interest occurs when the phases ϕ are uniformly distributed over the interval $(0, 2\pi)$; all moments $\langle\eta^k\rangle$ $(k\geq 1)$ vanish, and we are left with a spectral intensity $\langle a^2\rangle\langle S_k(f, r)S_l(f, r)^*\rangle/T_0$. However, when $\langle\eta\rangle \exp(i\omega T_0)=1$, one obtains besides the continuum (derived from Eq. (4.18)), an infinite number of periodic components (for details see the Appendix[9]). The only cases of importance occur for $\langle\eta\rangle=1$, $f=m/T_0$ or $\langle\eta\rangle=-1$, $f=(2m+1)/2T_0$. In the former instance, $[w_2(\phi)=\delta(\phi-0)]$, one also obtains a dc component. The cross-spectral density is explicitly

$$W_{kl}(f)=(2/T_0)\Big\{\langle a^2\rangle\langle S_k(f, r)S_l(f, r)^*\rangle -\langle a\rangle^2\langle c_k(f, r, \epsilon)\rangle\langle c_l(f, r, \epsilon)^*\rangle +\frac{\langle a\rangle^2}{T_0}\langle c_k(f, r, \epsilon)\rangle \times\langle c_l(f, r, \epsilon)^*\rangle\sum_{m=0}^{\infty}\delta\Big(\frac{(f-m)}{T_0}\Big)\Big\}, \quad (4.19)$$

where we have written

$$c_{k,l}(f, r, \epsilon)=S_{k,l}(f, r)\exp(-i\omega\epsilon).$$

In the latter case, which represents a train of pulses alternating in sign, $w_2(\phi)=\delta(\phi-\pi)$, we have Eq. (4.19) except that now $\delta(f-m/T_0)$ is replaced by $\delta(f-(2m+1)/2T_0)$, see Appendix.[9] As expected, the new period is now double the old period, and because the pattern on the average alternates in sign, all even harmonics (of $f_0/2$) are missing.

We consider next

(II) *Non-overlapping, nonperiodic impulse noise*:[10] Here a procedure similar to the preceding case is followed, except that now the durations of the elementary pulses e_k, e_l are determined by the interval between the random initiation times of successive pulses; consequently, we can expect no periodicities. The first moments (for the time $t_{j'}=t_0$) are obtained from the first derivatives of the characteristic function (see Eq. (2.19)) on applying Eqs. (3.4) and (3.11):

$$\mu_k^{(1)}=\nu_k(t_0)=\sum_{N=0}^{\infty}W_1(N)_T \times\sum_{m=0}^{N}\langle a_m\cos\psi_m e_k(t_0-t_m'; r_m)\rangle =\sum_{N=0}^{\infty}W_1(N)_T\sum_{m=0}^{N}\Big[a_m\cos\psi_m \times\int_{-\infty}^{\infty}S_k(f, r_m)\exp(i\omega t_0-i\omega t_m')df\Big]_{\text{stat av}}, \quad (4.20)$$

where Eq. (4.2) has been used. Considering next the stationary case, we have $T\to\infty$. Then the average number $\langle N\rangle$ of events in $(0, T)$ also becomes infinite. In the limit, then, $W_1(N)_T$ may safely be replaced by a continuous distribution density $W_1(N)_\infty dN$, where the maximum probability and therefore the maximum probability density is centered about $\langle N\rangle$ and in the immediate vicinity $[0(\langle N\rangle^{\frac{1}{2}})]$ of $\langle N\rangle$, the error in the replacement being $0(1/N)$. Accordingly, we have

$$\sum_{N=0}^{\infty}W_1(N)_{T\to\infty}\longrightarrow\int_0^{\infty}W_1(N)_\infty dN=1.$$

Furthermore, the number of intervals in $(0, T)$ having lengths between r and $r+dr$ is $(N+1)w(r)dr$, and their aggregate length is $r(N+1)w(r)dr$, so that the total length T is

$$T=\int_0^{\infty}r(N+1)w(r)dr=\langle r\rangle(N+1).$$

[10] For some simple examples, see G. W. Kenrick, Phil. Mag. **7**, 176 (1929), Ser. 7; S. O. Rice, Bell System Tech. J. **23**, 282 (1944), Sec. 2.7.

With these modifications, exchanging time and ensemble averages gives finally for the first moments

$$\nu_k=\frac{1}{\langle r\rangle}\lim_{N\to\infty}\frac{1}{N+1}\sum_{m=0}^{N}a_m\cos\psi_m S_k(0,r_m)$$

$$=\frac{\langle a\cos\psi S_k(0,r)\rangle}{\langle r\rangle};\quad(4.21)$$

see Eqs. (4.4a), (4.4b). The density parameter γ characteristic of this type of noise is always unity, for here $\gamma_{II}=(N+1)\langle r\rangle/T=1$. For rectangular pulses ν_k becomes simply $\langle a\cos\psi\rangle$.

The second moments are found in a similar way: one has Eq. (4.6), except that the exponential is replaced by $\exp[-i\omega t-i\omega(t_m'-t_n')]$ and T_0 by $\langle r\rangle$. Rearranging the series to sum over $m=m$ and $n=m+j$ and noting that now

$$t_{m+j}'-t_m'=r_m+d_{m,j}(r_\lambda),$$

where

$$d_{m,j}(r_\lambda)=\sum_{m+1}^{m+j-1}r_\lambda,\quad j\geq 2;\quad=0,\quad j=1;$$

$$=-r_m,\quad j=0;\quad=-\sum_{m}^{m+j}r_\lambda,\quad j\leq -1,$$

we can write, finally,

$$M_{kl}=\mu_{kl}^{(2)}=R_{kl}(t)$$

$$=\frac{1}{\langle r\rangle}\sum_{j=-\infty}^{\infty}\lim_{N\to\infty}\frac{1}{N+1}\sum_{m=0}^{N}a_m a_{m+j}\cos\psi_m\cos\psi_{m+j}$$

$$\times\int_{-\infty}^{\infty}df[S_k(f,r_m)\exp(i\omega r_m)][S_l(f,r_{m+j})^*]$$

$$\times\exp(-i\omega t-i\omega[d_{m,j}(r_\lambda)]).\quad(4.22)$$

Because the process is stationary, the time averages may be replaced by an equivalent joint statistical average in the manner of Eq. (4.9).

When the phases are cumulative, so that $\psi_m=\sum_{\lambda,m}\phi_\lambda$, and r, a, ϕ_λ, and $t_{(m)}'$ are all independent from pulse to pulse, the cross-correlation functions (4.22) simplify greatly. Writing

$$\alpha_m^{(k)}(\omega,\mathbf{a})=a_m S_k(f,r_m)\exp(i\omega r_m);$$

$$\beta_{m+j}^{(l)}(\omega,\mathbf{a})=a_{m+j}S_l(f,r_{m+j}),\quad(4.23)$$

and letting $z(\omega)=\exp(i\omega r)$, so that

$$\langle z(\omega)\rangle=\int_0^\infty w(r)\exp(i\omega r)dr,$$

we again replace time by ensemble averages in Eq. (4.22) and obtain for $j(\geq 1)$ a number of geometric series, like Eq. (4.16), which are directly summed if we exclude the case $\langle\eta\rangle\langle\exp(i\omega r)\rangle=1$, etc.; see Eqs. (4.15)–(4.17). Repeating for $(j\leq 0)$ and combining the results, we have finally the cross-correlation function

$$R_{kl}(t)=(1/4\langle r\rangle)$$

$$\times\int_{-\infty}^{\infty}\exp(-i\omega t)df\Big\{(\langle\eta^2\rangle+\langle\eta^2\rangle^*+2)$$

$$\times\langle\exp(-i\omega r)\alpha_1^{(k)}(\omega,\mathbf{a})\beta_1^{(l)}(\omega,\mathbf{a})^*\rangle$$

$$+\langle\alpha_1^{(k)}(\omega,\mathbf{a})\rangle\langle\beta_1^{(l)}(\omega,\mathbf{a})^*\rangle$$

$$\times\left[\frac{\langle\eta\rangle(\langle\eta^2\rangle+1)}{1-\langle z(\omega)\rangle\langle\eta\rangle}+\frac{\langle\eta^*\rangle(\langle\eta^2\rangle^*+1)}{1-\langle z(\omega)\rangle\langle\eta\rangle^*}\right]$$

$$+\langle\alpha_1^{(l)}(\omega,\mathbf{a})^*\rangle\langle\beta_1^{(k)}(\omega,\mathbf{a})\rangle$$

$$\times\left[\frac{\langle\eta\rangle(\langle\eta^2\rangle+1)}{1-\langle z(\omega)^*\rangle\langle\eta\rangle}+\frac{\langle\eta^*\rangle(\langle\eta^2\rangle^*+1)}{1-\langle z(\omega)^*\rangle\langle\eta^*\rangle}\right]\Big\}.\quad(4.24)$$

There is, in general, no dc component when $\langle z(\omega)\rangle\langle\eta\rangle$ (and consequently $\langle z(\omega)\rangle^*\langle\eta\rangle$, etc.) is different from unity. However, when $\langle\eta\rangle$ is unity, one finds for $f=0$ that $\langle\eta\rangle\langle z(\omega)\rangle$ is also unity; and the sum over j is seen (see Appendix[9]) to be proportional to $\delta(f-0)$, so that a constant term must be added to $R_{kl}(t)$, which is just the dc contribution $\nu_k\nu_l$. The cross-spectral intensity follows as before from Eq. (4.24) with the help of Eq. (3.17). Note that for purely random phases $w(\phi)=1/2\pi$, $(0\leq\phi<2\pi)$, all moments $\langle\eta^k\rangle$ vanish once again, leaving us with $\langle a^2\rangle\langle S_k(f,r)S_l(f,r)^*\rangle/\langle r\rangle$ as the spectral intensity. A number of special cases of spectra $(k=l)$ are given in reference 2, p. 45.

We have finally

(III) *Poisson noise*: Because of the independent, overlapping character of the elementary pulses, the explicit form of the characteristic function (2.28) is much simpler than for the nonoverlapping models (I) and (II). It is also now more natural to consider first the semi-invariants, and from them to determine the moments of interest according to Eqs. (3.2) and (3.3). Comparing Eq. (2.28) and the definining relation (3.3) one has at once for the second-order $(l=2)$, stationary process $(s=2n)$, the general Mth-order semi-invariant $(\gamma=\gamma_{III})$:

$$\lambda_{m(1)\cdots m(2n)}^{(M)}=$$

$$\gamma_{III}\left\{a^M\int_{-\infty}^{\infty}[\prod_{j=1}^{n}h_j(x,r)^{m(j)}h_j(x+\beta t,r)^{m(n+j)}]dx\right\}_{\text{stat av}}.$$

(4.25)

From Eq. (3.4) one finds that the averages are

$$\nu_k=\lambda_{0\cdots1\cdots0}{}^{(1)}$$

$$=\gamma_{III}\left[a\int_{-\infty}^{\infty}h_k(x,r)dx\right]_{\text{stat av}}$$

$$=\gamma_{III}\left[a\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}S_k(f,r)\exp(i\omega x/\beta)df\right]_{\text{stat av}}$$

$$=\gamma_{III}\langle a\beta S_k(0,r)\rangle, \tag{4.26}$$

where

$$S_k(f,r)=\beta^{-1}\int_{-\infty}^{\infty}dx\exp(i\omega x/\beta)h_k(x,r)$$

is the amplitude distribution of a typical pulse; see Eqs. (4.4) and (4.12). The second moments are also readily obtained in the same fashion, from Eqs. (3.7), (3.8), and (4.25). We have $m(j)=1$, $1\leq k=j\leq n$; $m(j+n)=1$, $1\leq l=j\leq n$, and so

$$\sigma_{kl}=\lambda_{(k,l)}{}^{(2)}$$

$$=\gamma_{III}\left[a^2\int_{-\infty}^{\infty}h_k(x,r)h_l(x+\beta t,r)dx\right]_{\text{stat av}} \tag{4.27}$$

giving the cross-correlation function

$$R_{kl}(t)=\gamma_{III}\left[a^2\beta\int_{-\infty}^{\infty}\times\exp(-i\omega t)S_k(f,r)S_l(f,r)^*df\right]_{\text{stat av}}$$

$$+\gamma_{III}{}^2\langle a\beta S_k(0,f)\rangle\langle a\beta S_l(0,f)^*\rangle. \tag{4.28}$$

The cross-spectral density follows immediately from Eq. (3.17) and is

$$W_{kl}(f)=2\gamma_{III}\langle\beta a^2S_k(f,r)S_l(f,r)^*\rangle$$

$$+2\gamma_{III}{}^2\beta^2\langle aS_k(0,f)\rangle\langle aS_l(0,f)^*\rangle\delta(f-0), \tag{4.29}$$

which is the generalization of an earlier result obtained by Rice.[11]

From Eqs. (2.37) and (2.38) with the help of Eq. (3.9), the various moments $\mu_{q(1)\cdots q(s)}{}^{(Q)}$ of the distribution are found from

$$\mu_{q(1)\cdots q(s)}{}^{(Q)}=\langle X_1{}^{q(1)}\cdots X_s{}^{q(s)}\rangle$$

$$=(-i)^Q\left\{\frac{\partial^Q}{\partial\xi_1{}^{q(1)}\cdots\partial\xi_s{}^{q(s)}}\sum_{n=1}^{\infty}\frac{\gamma_{III}{}^n}{n!}\times\Phi_l(\xi_1,t_1;\cdots;\xi_st_s)\right\}\Bigg|_{\xi_1=\xi_2=\cdots=\xi_s=0}, \quad Q\geq1. \tag{4.30}$$

When $Q=0$, one gets the trivial result unity. We observe that when n exceeds Q, every term in the braces $\{\ \}$ vanishes, since each such term contains a factor Φ^{n-Q} at least, which is zero; see Eq. (2.36). The Qth moment is thus a polynomial in γ_{III}, whose term of highest power is $\gamma_{III}{}^Q$. Therefore, in the case of impulsive random noise it is necessary in order to determine the first Q moments, to obtain the series development of the corresponding probability density $W_l(2.37)$ through the term in $\gamma_{III}{}^Q$ only. No additional terms in γ_{III}, or moments beyond the Qth order are needed, unlike the case of nearly normal random noise (Sec. 5), where in principle all the (inverse) powers of γ_{III} are required to specify a given Qth moment.

(ii) Narrow-Band Waves

As in (i) above, we want explicitly the first and second moments from the second-order probability density for the $2n$ stationary random variables X_j $(j=1, 2\cdots2n)$, where now we consider narrow-band disturbances of the form (2.6b) and noise models I–III:

(I) Following the approach of (i), we find for the periodic wave trains that the average can be expressed as

$$\nu_k(t_0)=\sum_{m=0}^{N}\left\{a_m\int_{-\infty}^{\infty}\exp[i\omega(mT_0+\epsilon_m)+i\omega t_0]\times\{\exp(i\psi_m)S_k(f-f_c,r_m)+\exp(-i\psi_m)S_k(f+f_c,r_m)\}df/2\right\}_{\text{stat av}}, \tag{4.31}$$

where $S_k(f-f_c,r_m)$ is given by Eq. (4.2), and f_c is the carrier frequency of the pulsed oscillations. In the stationary cases the initial time (t_0) drops out, and we have finally, corresponding to (4.4),

$$\nu_k=(1/T_0)\mathcal{R}e\langle ae^{i\psi}S_k(f_c,r)^*\rangle. \tag{4.32}$$

Because of the narrow-band character of the elementary pulses, $S_k(f-f_c,r)$ is significant only for frequencies f in the immediate spectral vicinity of the carrier f_c, so that in most cases the mean value ν_k may be safely ignored. However, the fact that there is in general a finite dc component is apparent if we observe now that the pulse train is really a train of transient oscillations, whose amplitude usually decreases with duration, leaving a net steady contribution.

The second moments follow as before [see Eqs. (4.5)-(4.13)], except that in place of $A_m{}^{(k)}(\omega,\mathbf{a})$, $A_{m+j}{}^{(l)}(\omega,\mathbf{a})^*$, Eq. (4.7), one writes instead

$$B_m{}^{(k)}(\omega-\omega_c,\mathbf{a})=a_m\exp(-i\omega\epsilon_m)\times[\exp(i\psi_m)S_k(f-f_c,r_n)+\exp(-i\psi_m)S_k(f+f_c,r_m)]/2;$$

$$B_{m+j}{}^{(l)}(\omega-\omega_c,\mathbf{a})^*=a_{m+j}\exp(i\omega\epsilon_{m+j})\times[\exp(-i\psi_{m+j})S_l(f-f_c,r_{m+j})^*+\exp(i\psi_{m+j})S_l(f+f_c,r_{m+j})^*]/2. \tag{4.33}$$

[11] S. O. Rice, Bell System Tech. J. **23**, 282 (1944), Eq. (2·6-4).

Our previous relations (4.8)–(4.13) still apply with this modification, except that Eq. (4.11) is now written

$$R_{kl}(t)\doteq\frac{1}{T_0}\sum_{j=-\infty}^{\infty}\langle a_1a_2\cos[\omega_c(t-jT_0+\epsilon_1-\epsilon_2)$$
$$+\psi_2-\psi_1]\rho_{kl}(t-jT_0+\epsilon_1-\epsilon_2;r_1,r_2)_0\rangle, \quad (4.34)$$

where $\rho_{kl}(\ )_0$, Eq. (4.10), is now the (normalized) cross-correlation function for the slowly varying parts of the waves. We have applied the narrow-band condition $S_{k,l}(f+f_c,r)=0$ to discard terms involving $\cos[2\omega_0(t_0+t-jT_0+\epsilon_1-\epsilon_2)+\psi_1-\psi_2]$ in the integrand of $R_{kl}(t)$ [see Eq. (4.8)]. The factor 2 which appears in Eq. (4.11) is suppressed, because we are now dealing with *peak* values a, e_k, etc., a factor 2^{-1} appearing here from the average of carrier terms of the type $\cos\omega_c t_0\cos\omega_c(t_0+t)$. The example given in Eq. (4.14) reduces in this instance to

$$W_{kl}(f)\doteq\frac{S_k(f-f_c,r_0)S_l(f-f_c,r_0)^*}{2T_0}$$
$$\times\left\{\langle a^2\rangle-\langle a\rangle^2|\,b(f)\langle e^{i\psi}\rangle|^2\right.$$
$$\left.+\frac{\langle a\rangle^2}{T_0}|\,b(f)\langle e^{i\psi}\rangle|^2\sum_{m=0}^{\infty}\delta\left(\frac{(f-m)}{T_0}\right)\right\}. \quad (4.35)$$

When the phase is cumulative, i.e., $\psi_m=\sum_{\lambda=0,m}\phi_\lambda$, and when a, r, ϵ, and ϕ_λ are independent from period to period, one finds by a straightforward modification of Eq. (4.18) that

$$R_{kl}(t)\doteq(1/4T_0)\int_{-\infty}^{\infty}\exp(-i\omega t)df$$
$$\times\{\langle C_1^{(k)}(\omega-\omega_c,\mathbf{a})C_1^{(l)}(\omega-\omega_c,\mathbf{a})^*\rangle$$
$$+2\Re e\langle C_1^{(k)}(\omega-\omega_c,\mathbf{a})\rangle\langle C_1^{(l)}(\omega-\omega_c,\mathbf{a})^*\rangle$$
$$\times[\langle\eta^*\rangle\exp(+i\omega T_0)/[1-\langle\eta^*\rangle\exp(i\omega T_0)]]\}. \quad (4.36)$$

The cross-correlation functions $R_{kl}^{(yy)}(t)$, $R_{kl}^{(yz)}(t)$, $R_{kl}^{(zz)}(t)$, [see Sec. 4, (ii)] for the slowly varying parts of the noise waves X_k, X_l are given by Eqs. (4.8) and (4.9) on replacing $a_m\cos\psi_m$ and $a_{m+j}\cos\psi_{m+j}$ in $A_m^{(k)}$, $A_{m+j}^{(l)}$ by $a_ma_{m+j}\cos[\omega_c(mT_0+\epsilon_m)-\psi_m]\times\cos([\omega_c(m+j)T_0+\epsilon_{m+j}]-\psi_{m+j})$, $a_ma_{m+j}\cos(\ )_m\times\sin(\ )_{m+j}$, and $a_ma_{m+j}\sin(\ )_m\sin(\ )_{m+j}$, respectively.

(II) Here for the average ν_k one has Eq. (4.32) on replacing T_0 by the mean duration $\langle r\rangle$. The second moments are also easily found for the narrow-band case, giving us finally for the cross-correlation function (4.24) the somewhat simpler version,

$$R_{kl}(t)\doteq\frac{1}{4\langle r\rangle}\int_{-\infty}^{\infty}\exp(-k\omega t)df$$
$$\times\{\langle\exp(-i\omega r)\alpha_1^{(k)}(\omega-\omega_c,\mathbf{a})\beta_1^{(l)}(\omega-\omega_c,\mathbf{a})^*\rangle$$
$$+[\langle\alpha_1^{(k)}(\omega-\omega_c,\mathbf{a})\rangle\langle\beta_1^{(l)}(\omega-\omega_c,\mathbf{a})^*\rangle$$
$$\times\langle\eta^*\rangle/[1-\langle\eta^*\rangle\langle z(\omega)\rangle]]+\langle\alpha_1^{(l)}(\omega-\omega_c,\mathbf{a})^*\rangle$$
$$\times\langle\beta_1^{(k)}(\omega-\omega_c,\mathbf{a})\rangle\langle\eta\rangle/[1-\langle\eta\rangle\langle z(\omega)\rangle^*]\}. \quad (4.37)$$

The frequency f_c now represents the central frequency of the noise band. Observe that even when $\langle\eta\rangle=1$ (i.e., $w_2(\phi)=\delta(\phi-0)$), there are no discrete components at f_c, $2f_c$, etc., because the *duration* of the oscillations is random, each segment r seconds long ends at a different time (i.e., phase) on the average. Correspondingly, if $w_1(r)=\delta(r-r_0)$, all segments being of the same length, one does have periodicities, which appear when $\langle\eta\rangle\langle z(\omega)\rangle$ is unity, as above. The low frequency cross-correlation functions $R_{kl}^{(yy,yz,zz)}(t)$ are given by

$$R_{kl}^{(yy,yz,zz)}(t)$$
$$\doteq\frac{1}{\langle r\rangle}\sum_{j=-\infty}^{\infty}\int_{-\infty}^{\infty}dfR_{kl}^{(yy,yz,zz)}(j;\omega)_d\exp(-i\omega t), \quad (4.38)$$

where now $R_{kl}^{(yy)}(j,\omega)_d$ is found by replacing $a_m\cos\psi_m\times\exp(-i\omega\epsilon_m)$ in $A_m^{(k)}$ by $a_m\cos(\omega_c t_m'-\psi_m)$ and $a_{m+j}\times\cos\psi_{m+j}\exp(-i\omega\epsilon_{m+j})$ in $A_{m+j}^{(l)}$ with $a_{m+j}\cos(\omega_c t_{m+j}'-\epsilon_{m+j})$, and $S_k(f,r)$ by $S_k(f-f_c,r)$, etc. For $R_{kl}^{(yz)}(j,\omega)_d$ and $R_{kl}^{(zz)}(j,\omega)_d$ one proceeds in a similar way getting the following combinations: $\cos(\omega_c t_m'-\psi_m)\sin(\omega_c t_{m+j}'-\psi_{m+j})$ and $\sin(\omega_c t_m'-\psi_m)\sin(\omega_c t_{m+j}'-\psi_{m+j})$, respectively.

(III) Finally, for poisson noise one has the average $\nu_k=\gamma_{III}\Re e\langle a\exp(i\psi)S_k(-f_c,r)\rangle$, which is essentially zero for most practical purposes. The cross-correlation function (4.28) becomes now

$$R_{kl}(t)\doteq\tfrac{1}{2}\gamma_{III}\left\{a^2\cos[\omega_c t+\psi_l-\psi_k]\right.$$
$$\left.\times\int_{-\infty}^{\infty}h_k(x,r)h_l(x+\beta t,r)dx\right\}_{\text{stat av}}+\nu_k\nu_l, \quad (4.39)$$

where we have neglected the term in the integrand containing $\cos(2\omega_c x/\beta)$, etc., because of the narrow-band character of our wave. Here ν_k and ν_l are determined by the systems through which the original waves have passed. Again the factor $\frac{1}{2}$ which appears in (4.39), but not in (4.28) for broad-band disturbances, is easily explained if we remember that h_k, h_l are essentially peak values. In the former case h_k and h_l are elementary envelopes [see (2.6b)] and consequently the mean

power ($R_{kk}(0)$) and correlation functions will contain the factor $\frac{1}{2}$, corresponding to the average over a cycle of the "carrier" or central frequency component $\cos(\omega_c t-\theta)$. In the latter, the factor $\frac{1}{2}$ is replaced by unity, since there is no "carrier" term. Setting $\omega_c=0$ and including in the above terms involving $\cos 2\omega_c x/\beta$, etc., do yield unity in (4.28). The corresponding spectral density for the narrow-band waves is

$$W_{kl}(f)=\frac{\beta}{2}\gamma_{III}\langle a^2\exp[-i(\psi_l-\psi_k)]\times S_k(f-f_c,r)S_l(f-f_c,r)^*\rangle+2\nu_k\nu_l\delta(f-0). \quad (4.40)$$

The various low frequency correlation functions $R_{kl}^{(\nu\nu)}(t)$, etc. follow in similar fashion.

5. NEARLY NORMAL RANDOM NOISE

In the case of poisson noise, when the number of elementary random impulses is large and their mean duration is not too brief (a condition expressed by the relation $\gamma_{III}\gg 1$) heavy overlapping of the pulses occurs, and one has then *nearly normal random noise*. To determine the explicit form of the second-order probability density W_2 in this instance the characteristic function $F_2(\xi_1\cdots\xi_{2n};t)$ is first given by a development in the semi-invariants (3.3), (4.25) of the distribution. The exponent of (3.3) is separated into a quadratic form and a part containing higher powers of $\xi_1\cdots\xi_{2n}$. As is well known,[6,11-13] the first exponential yields a multivariate normal distribution. The correction terms, whose contributions cannot be ignored for certain types of precipitation and clutter (see Table I) even for $\gamma_{III}\gg 1$, may be found by expanding[14] the exponential with the higher powers of $\xi_1\cdots\xi_{2n}$ and collecting terms in such a way as to give a development of the corresponding probability density according to inverse powers of γ_{III}. Noting that

$$F_2(\xi_1\cdots\xi_{2n};t)=\exp(i\mathbf{v}'\boldsymbol{\xi}-\tfrac{1}{2}\boldsymbol{\xi}'\boldsymbol{\sigma}\boldsymbol{\xi})\times[1+E_2(\xi_1\cdots\xi_{2n};t)], \quad (5.1)$$

where E_2 is the desired correction term, we obtain the probability density as the fourier transform of (5.1), the leading term of which yields the multivariate normal distribution

$$W_2(X_1\cdots X_{2n};t)_\infty=((2\pi)^n|\sigma|^{\frac{1}{2}})^{-1}\times\exp[-\tfrac{1}{2}(\mathbf{X}'-\mathbf{v}')\mathbf{A}(\mathbf{X}-\mathbf{v})]. \quad (5.2)$$

The matrix $\mathbf{A}$ has the elements $\sigma^{kl}/|\sigma|$; σ^{kl} is the cofactor of the element σ_{kl} in $\boldsymbol{\sigma}$, and $|\sigma|$ is the determinant of $\boldsymbol{\sigma}$. The correction terms for $W_2(X_1\cdots X_{2n};t)$ require integrals of the form,

$$I_S(X_1\cdots X_{2n};t)=\int_{-\infty}^{\infty}\frac{d\xi_1}{2\pi}\cdots\int_{-\infty}^{\infty}\frac{d\xi_{2n}}{2\pi}\xi_1^{s(1)}\cdots\xi_{2m}^{s(2m)}\times\exp[-i\boldsymbol{\xi}'(\mathbf{X}-\mathbf{v})-\tfrac{1}{2}\boldsymbol{\xi}'\boldsymbol{\sigma}\boldsymbol{\xi}].$$
$$(S=s(1)+\cdots+s(2m)\geq 0). \quad (5.3)$$

With the aid of the normalizing transformations $\boldsymbol{\xi}=\mathbf{a}^{-1}\boldsymbol{\zeta}$, $\mathbf{y}=\mathbf{a}'^{-1}(\mathbf{X}-\mathbf{v})$, where $\mathbf{a}$ is the diagonal matrix $\|(\sigma_{ij})^{\frac{1}{2}}\delta_{ij}\|$, so that $\mathbf{a}'=\mathbf{a}$ and $\mathbf{a}'^{-1}=\mathbf{a}^{-1}$, we have finally

$$I_S=i^S\sigma^{-s(1)/2}\cdots\sigma_{2n,2n}^{-s(2n)/2}\frac{\partial^S}{\partial y_1^{s(1)}\cdots\partial y_{2m}^{s(2m)}}\times W_2(y_1\cdots y_{2n};t)_\infty$$
$$=i^S\sigma^{-s(1)/2}\cdots\sigma_{2n,2n}^{-s(2n)/2}\frac{\partial^S}{\partial y_1^{s(1)}\cdots\partial y_{2m}^{s(2m)}}\times(2\pi)^{-n}|\sigma|^{-\frac{1}{2}}\exp(-\mathbf{y}'\boldsymbol{\sigma}''\mathbf{y}/2) \quad (5.4)$$

with $\boldsymbol{\sigma}''=\mathbf{a}^{-1}\boldsymbol{\sigma}\mathbf{a}'^{-1}=\|\sigma_{kl}/(\sigma_{kk}\sigma_{ll})^{\frac{1}{2}}\|$, $|a|=\det\mathbf{a}$. The matrix $\boldsymbol{\sigma}''$ is accordingly independent of γ_{III} and $y_j[=(X_j-\nu_j)/(\sigma_{jj})^{\frac{1}{2}}]$, $(j=1\cdots 2n)$, is dimensionless. The normal distribution density $W_2(y_1\cdots y_{2n};t)_\infty$ in the reduced variables y_j is found directly from (5.2) with the help of the transformation $\mathbf{y}=\mathbf{a}'^{-1}(\mathbf{X}-\mathbf{v})$.

Since each element of $\mathfrak{A}^{(M)}$ and hence all σ_{jj}, are proportional to γ_{III} [see (4.26)], we may now expand E_2 and use Eq. (5.4) to arrange $W_2(X_1\cdots X_{2n};t)$ in descending powers of γ_{III}. The leading term follows immediately from (5.4) on setting $S=0$ and is given by (5.2). The dependence on γ_{III} is $O(\gamma_{III}^{-n})$, from the factor $|\sigma|^{-\frac{1}{2}}$. The correction terms are easily found if we observe that whenever $\xi_j^{s(j)}$ appears in the integrand of (5.3) a factor $\sigma_{jj}^{-s(j)/2}$ also appears. We write finally

$$E_2(\xi_1\cdots\xi_{2m};t)=\sum_{M=3}^{4}\frac{(i\xi_1)^{m(1)}\cdots(i\xi_{2n})^{m(2n)}}{m(1)!\cdots m(2n)!}\lambda_{m(1)\cdots m(2n)}^{(M)}+\frac{1}{2!}\left[\sum_{M=3}^{3}\frac{(i\xi_1)^{m(1)}\cdots(i\xi_{2n})^{m(2n)}}{m(1)!\cdots m(2n)!}\lambda_{m(1)\cdots m(2n)}^{(3)}\right]^1+O_T(\gamma_{III}^{-\frac{1}{2}}). \quad (5.5)$$

The symbol $O_T(\gamma_{III}^{-\frac{1}{2}})$ indicates that when the fourier transform of E_2 is taken, the remainder is $O(\gamma_{III}^{-\frac{1}{2}})$. For $M=3$ we observe in W_2 a factor $\gamma_{III}^{-n-\frac{1}{2}}$; for $M=4$ and from the squared-series in (5.5) above we obtain γ_{III}^{-n-1} for the dependence on γ_{III}. Higher order corrections follow in a similar manner. The general

[12] A. Blanc-Lapierre, dissertation, University of Paris (Masson et Cie, Paris, 1945).
[13] V. Boonimovitch, J. Phys. U.S.S.R. **10**, 35 (1946).
[14] Rigorous demonstration of the asymptotic properties of series of this type has been given by H. Cramér, *Random Variables and Probability Distributions*, Cambridge Tract. in Math. No. 36 (1937).

characteristic function (5.1) is therefore

$$F_2(\xi_1\cdots\xi_{2n};t)=\exp(i\mathbf{v}'\boldsymbol{\xi}-\tfrac{1}{2}\boldsymbol{\xi}'\boldsymbol{\sigma}\boldsymbol{\xi})\times\left\{1+\sum_{M=3}^{4}b_M+\frac{1}{2!}(\sum_{M=3}^{3}b_M)^2+O_T(\gamma_{III}^{-\frac{1}{2}})\right\},\quad(5.6)$$

where b_M represents the summand in (5.5). The associated probability density has the asymptotic form,

$$W_2(X_1\cdots X_{2n};t)\simeq[(2\pi)^n|\sigma|^{\frac{1}{2}}]^{-1}\times\left\{1-\sum_{M=3}^{3}\mathbf{C}_M+\sum_{4}^{4}\mathbf{C}_M+\frac{1}{2!}(\sum_{3}^{3}\mathbf{C}_3)^2+O(\gamma_{III}^{-\frac{1}{2}})\right\}\exp(-\tfrac{1}{2}\mathbf{y}'\boldsymbol{\sigma}''\mathbf{y}),\quad(5.7)$$

in which

$$\mathbf{C}_M=[\lambda_{m(1)\cdots m(2n)}{}^{(M)}/m(1)!\cdots m(2n)!\sigma_{11}{}^{m(1)/2}\cdots\sigma_{2n,2n}{}^{m(2n)/2}]\partial^M/\partial y_1{}^{m(1)}\cdots\partial y_{2n}{}^{m(2n)};$$

$$\mathbf{y}=\mathbf{a}'^{-1}(\mathbf{X}-\mathbf{v}).\quad(5.8)$$

In the one-dimensional case (5.7) is known as Edgeworth's series.[4,11] The first-order density $W_1(X_1\cdots X_n)$ is found immediately from the above formally by restricting the total number of random variables to $X_1\cdots X_n$ only, with appropriate modifications in $\mathbf{v}$, $\boldsymbol{\sigma}$, etc.; a similar adjustment is made for the characteristic function $F_1(\xi_1\cdots\xi_n)$. Note that in the limit $\gamma_{III}\to\infty$ one readily obtains the predicted normal distribution (5.2), since the correction terms vanish more rapidly than the leading term. This is a special case of the Central Limit Theorem of probability,[15] for which the sufficiency conditions in the present cases are easily shown to apply.

6. CONCLUDING REMARKS

For noise models I and II the explicit evaluation of the probability densities W_1, W_2, $\cdots W_n\cdots$, etc. is in general a very difficult task [see Eqs. (2.11), (2.12), (2.19)]. Only in relatively few special instances is the analytic form manageable. Poisson noise (III) [see Eqs. (2.23), (2.27), (2.28)], however, is more tractable. This occurs because one has now to consider completely independent, elementary effects, where the order of their occurrence is not important, unlike noise (I) and (II), which specifically requires a definite sequence. The overlapping, independent character of the elementary pulses here is also the reason for the normal distribution law when the density γ_{III} becomes extremely large. In the non-overlapping cases the density parameters γ_I, γ_{II} cannot exceed unity, and no asymptotic form of the distribution is then possible, of course. Table II gives the specific density parameters $\gamma(\equiv$[average number (n) of impulses/sec]$\times$[mean duration $(\langle\tau\rangle=\beta^{-1})$ of the pulses]) for the three types of noise considered in the present paper. The poisson noise model appears to embrace the widest variety of physical cases (see Table I),[1] although the periodic and nonperiodic models (I, II) are essential in communication theory for an effective discussion of pulse-time modulation schemes. It is fortunate, then, that even in nonlinear situations (see Sec. 1), when *all* moments of the (second-order) probability density are required (and here the second-order density itself), it is possible to obtain satisfactory approximations for low density poisson or impulsive noise, provided γ_{III} is $O(10^{-1})$, the usual case. Here pulse-shape is critical. The intermediate situations where γ_{III} is $O(10^0-10^1)$ are far more difficult because pulse shape is still critical and because of the slow convergence of the analytic expressions (see Sec. 2), but the higher densities $(\gamma_{III}\gg1)$ characteristic of precipitation, shot noise, etc., are simplest to treat. One has now a normal distribution, with or without correction terms (see Sec. 5), and pulse shape is no longer important (except in determining the high frequency cutoff, a spectral region not required in most practical problems.

TABLE II.

Type	γ	Range of γ
I. Periodic (T_0), non-overlapping noise	$\gamma_I=\langle\tau\rangle/T_0$	$0\leq\gamma_I\leq1$
II. Nonperiodic, non-overlapping noise	$\gamma_{II}=1$	
III. Poisson noise	$\gamma_{III}=n\langle\tau\rangle$	$0\leq\gamma_{III}<\infty$

[15] J. V. Uspensky, *Introduction to Mathematical Probability*, (McGraw-Hill Book Company, Inc., New York, 1937), Chapters 14 and 15.

Emission of Radiation from Diatomic Gases. II. Experimental Determination of Effective Average Absorption Coefficients of CO

S. S. PENNER AND D. WEBER
*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California**
(Received March 30, 1951)

Empirical evidence obtained for CO at room temperature, which supports the use of average absorption coefficients for entire vibration-rotation bands, is examined. In Sec. II the existence of an average apparent absorption coefficient for data obtained with inadequate spectral resolution is demonstrated. In Sec. III semiquantitative evidence at elevated total pressures and small optical densities for the use of a constant absorption coefficient equal to the ratio of integrated absorption to effective band width is described. These results serve as direct support for a proposed simplified method for radiant-heat-transfer calculations.

I. INTRODUCTION

IN a previous publication a simple approximate method for carrying out emissivity calculations on diatomic gases at elevated total pressures was proposed.[1] The approximate method involves estimating an effective band width for vibration-rotation bands as a function of temperature by examining the temperature dependence of the dominant term in the expression for the integrated absorption of the rotational lines composing a vibration-rotation band. This method for estimating approximate band widths is in agreement with room-temperature infrared-transmission measurements on carbon monoxide.[2]

The approximate method for emissivity calculations also involves the assumption that an average absorption coefficient, equal to the ratio of integrated absorption (for the entire vibration-rotation band) to effective band width, can be employed. It is the purpose of this discussion to present empirical data obtained for carbon monoxide at room temperature and elevated pressures in support of the proposed method for estimating average absorption coefficients.

II. APPARENT AVERAGE ABSORPTION COEFFICIENTS

Representative results of apparent absorption coefficients as a function of wavelength are shown in Fig. 1 for CO for slit widths of approximately 4 and 20 cm^{-1}, respectively. The apparent absorption coefficient for a slit width of 20 cm^{-1} is designated as P_ν', whereas the corresponding absorption coefficient for a slit of 4 cm^{-1} is P_ν''. For purposes of comparison a quantity proportional to the integrated absorption for the rotational transitions is plotted in Fig. 2. The ordinate in Fig. 2 is proportional to the integrated absorption per rotational line and has been calculated by using a theoretical relation obtained by Oppenheimer.[3–5] Wave numbers† in Figs. 1 and 2 were calculated from the assignments of Sponer.[6]

The experimental data shown in Fig. 1 can be explained by carrying through an extended analysis of

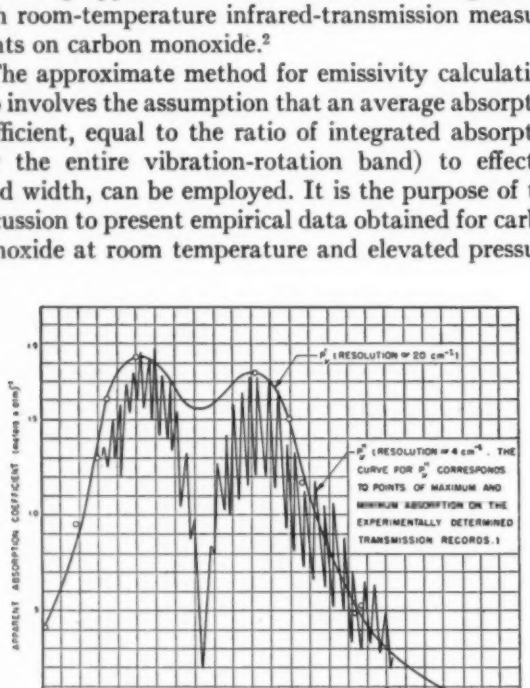

P_ν' (RESOLUTION ≈ 20 cm^{-1})
P_ν'' (RESOLUTION ≈ 4 cm^{-1}). THE CURVE FOR P_ν'' CORRESPONDS TO POINTS OF MAXIMUM AND MINIMUM ABSORPTION ON THE EXPERIMENTALLY DETERMINED TRANSMISSION RECORDS.)
APPARENT ABSORPTION COEFFICIENT (meters a atm)⁻¹
WAVELENGTH (microns)

FIG. 1. Apparent absorption coefficients as a function of wavelength for the fundamental vibration-rotation band of CO.

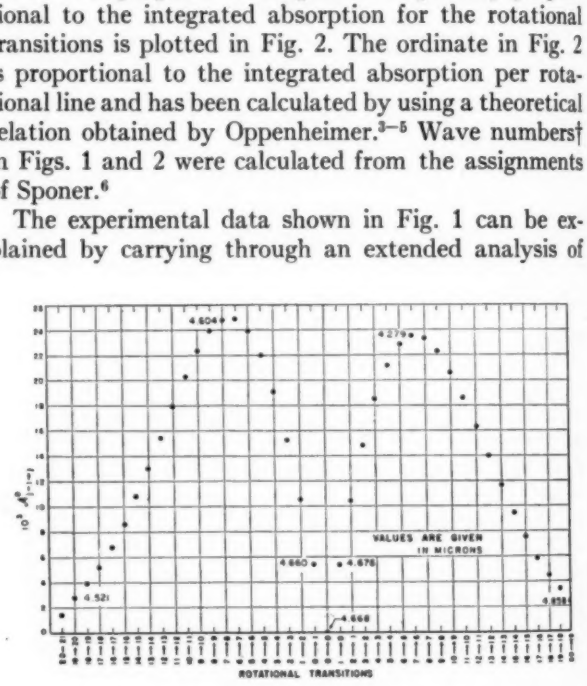


FIG. 2. Dependence of integrated absorption on rotational transition for the fundamental vibration-rotation band of CO at 298.7°K.

* This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. DA-04-495-ORD 18, sponsored by the U. S. Army Ordnance Department.

[1] S. S. Penner, J. Appl. Phys. **21**, 685 (1950); L. E. Benitez and S. S. Penner, *ibid.* **21**, 907 (1950).

[2] S. S. Penner, J. Appl. Mech. **18**, 53 (1951).

[3] J. R. Oppenheimer, Proc. Cambridge Phil. Soc. **23**, 327 (1926).

[4] D. M. Dennison, Phys. Rev. **31**, 503 (1928).

[5] S. S. Penner, J. Chem. Phys. **19**, 272 (1951).

† The wave number assignments for CO can be improved. However, the use of Sponer's values will introduce negligibly small errors into the present discussion. (See reference 5.)

[6] H. Sponer, *Molekülspektren* (Verlag. Julius Springer, Berlin, 1935).

the effect of slit width and prism resolution on experimental results.[7]

Experimental data of the type shown in Fig. 1 are, of course, not suitable for the determination of integrated absorption α.[8] The apparent integrated absorption α_F' for the fundamental of CO obtained from the data given in Fig. 1 is too small[8] by a factor of about 10. Nevertheless, as is shown in Table I, intensity ratios and relative positions of apparent absorption coefficients are seen to agree well with theoretical predictions.

Schack[9] concluded from an examination of absorption-coefficient data of the type given in Fig. 1 for P_ν' that, as a first approximation, the entire vibration-rotation band can be considered as being contained within a triangular envelope along each leg of which

$$P_\nu' = a + b\nu, \tag{1}$$

where a and b are constants. The total apparent intensity of radiation absorbed by the vibration-rotation band of width $\Delta\nu$ is

$$\int_{\Delta\nu} I_\nu d\nu = \int_{\Delta\nu} I_{o\nu}[1-\exp(-P_\nu' X)]d\nu, \tag{2}$$

where I_ν and $I_{o\nu}$ represent the absorbed and incident intensity, respectively, in the wave number interval between ν and $\nu+d\nu$ for an absorption cell containing CO at optical density X.

Setting $P_\nu'=0$ at $\nu=\nu_o$ and $P_\nu'=P_{\max}'$ at $\nu=\nu_{\max}$, it follows from Eqs. (1) and (2) that

$$\int_{\nu_o}^{\nu_{\max}} I_\nu d\nu = I_o(\nu_{\max}-\nu_o)\left[1-\frac{1-\exp(-P_{\max}'X)}{P_{\max}'X}\right] \tag{3}$$

TABLE I. Comparison of results obtained for the apparent absorption coefficients P_ν' and for theoretical values of integrated absorption.

	Calculated from data for P_ν' given in Fig. 1	Calculated from theoretical integrated absorption given in Fig. 2
Wavelengths at maxima	4.60 to 4.62 and 4.73μ	4.605 and 4.72μ
Ratio of maximum value of R branch to maximum value of P branch	1.04μ	1.055μ
Ratio of total intensity of R branch to total intensity of P branch	1.10μ	1.075μ
Wavelength below which negligible absorption should occur	~4.47μ	~4.47μ
Wavelength above which negligible absorption should occur	~4.95μ	~4.96μ

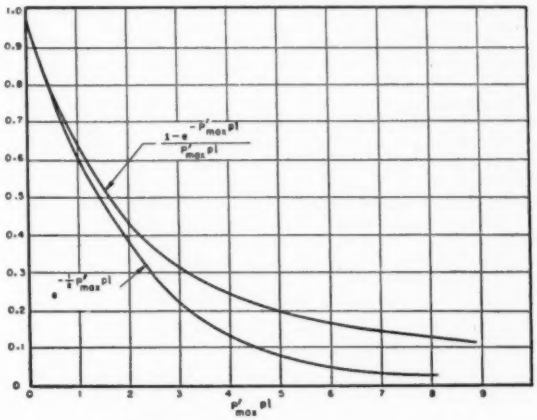


FIG. 3. The $\bar{P}'=P_{\max}'/2$ approximation if P_ν' is a linear function of ν.

if $I_{o\nu}=I_o$ is independent of ν. Equation (3) permits the calculation of the apparent absorbed radiant intensity provided $P_{\max}'$ and $\nu_{\max}-\nu_o$ are known.

The average apparent absorption coefficient $\bar{P}'$ for the frequency interval between ν_o and $\nu_{\max}$ is defined by the relation,

$$\int_{\nu_o}^{\nu_{\max}} I_\nu d\nu \simeq I_o(\nu_{\max}-\nu_o)[1-\exp(-\bar{P}'X)], \tag{4}$$

if the incident intensity of radiation is independent of wave number. Hence, if P_ν' depends on ν according to Eq. (1), a comparison of Eqs. (3) and (4) shows that $\bar{P}'$ can be expressed in terms of $P_{\max}'$ according to the relation,

$$\exp(-\bar{P}'pl) \simeq [1-\exp(P_{\max}'pl)]/P_{\max}'pl, \tag{5}$$

where the optical density X has been replaced by the product of pressure p and path length l. From Eq. (5) it is evident that

$$\bar{P}' \simeq \tfrac{1}{2}P_{\max}' \tag{6}$$

provided $P_{\max}'pl$ is much less than 1. The error inherent in the approximation of Eq. (6) can be ascertained by reference to Fig. 3, in which the function $\exp(-\bar{P}'pl)$, evaluated according to Eq. (5), is compared with $\exp(-\frac{1}{2}P_{\max}'pl)$ for various values of $P_{\max}'pl$. It should be noted that the preceding treatment is independent of the sign of a or b in Eq. (1) and that the relation $\bar{P}' \simeq \frac{1}{2}P_{\max}'$ represents an equally good approximation for each half and therefore for an entire triangle which nearly coincides with the envelope of an absorption band.

When measurements of percentage of transmission are made on an infrared spectrometer without using an automatic slit drive, the incident intensity $I_{o\nu}$ will, in general, be dependent on ν. For small frequency intervals it is justified to assume that $I_{o\nu}$ is a linear function of ν. If

$$I_{o\nu} = c + d\nu,$$

[7] Nielsen, Thornton, and Dale, Revs. Modern Phys. **16**, 307 (1944).

[8] S. S. Penner and D. Weber, J. Chem. Phys. **19**, 807, 817 (1951). See also the references cited in these articles.

[9] A. Schack, Z. techn. Physik **5**, 267 (1924).

TABLE II. Dependence of τ' for the fundamental of CO on pl at $p_T=700$ psia (CO pressurized with H_2, He, or A).

pl (cm atmos)	τ'	$\log\tau'$	Added gas
0.0203	0.9892	−0.0047	H_2
0.0339	0.9827	−0.0076	↓
0.0678	0.9433	−0.0253	↓
0.102	0.9259	−0.0334	↓
0.136	0.9044	−0.0436	↓
0.271	0.8200	−0.0862	↓
0.407	0.7563	−0.1213	↓
0.542	0.6996	−0.1551	↓
1.016	0.5959	−0.2248	↓
1.355	0.5578	−0.2535	↓
0.0203	0.9925	−0.0033	He
0.407	0.7537	−0.1228	↓
1.355	0.5415	−0.2664	↓
0.541	0.7152	−0.1456	A
0.678	0.6648	−0.1773	↓
1.355	0.5445	−0.2640	↓

where c and d are constants independent of ν, Eqs. (1) and (2) lead to the result:

$$\int_{\nu_o}^{\nu_{max}} I_\nu d\nu = c(\nu_{max}-\nu_o)\left[1-\frac{1-\exp(-P_{max}'X)}{P_{max}'X}\right] + d(\nu_{max}-\nu_o)\left\{\left[\frac{(\nu_{max}+\nu_o)}{2}\right]-\frac{\nu_{max}-\nu_o}{(P_{max}'X)^2} \times[1-\exp(-P_{max}'X)]-\frac{1}{P_{max}'X} \times[\nu_o-\nu_{max}\exp(-P_{max}'X)]\right\}, \quad (7)$$

whereas Eq. (4) should be replaced by the relation,

$$\int_{\nu_o}^{\nu_{max}} I_\nu d\nu \simeq (\nu_{max}-\nu_o) \times[c+d(\nu_{max}+\nu_o)/2][1-\exp(-\bar{P}'X)]. \quad (8)$$

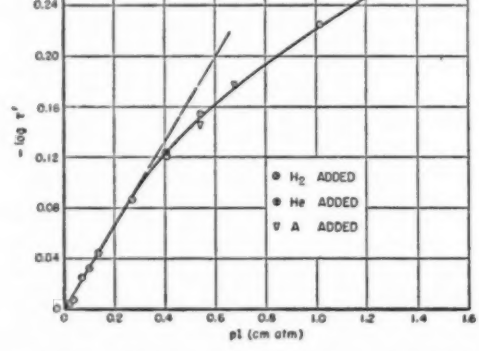


FIG. 4. Plot of $\log\tau'$ as a function of pl for the fundamental of CO (CO pressurized with H_2, He, or A to 700 psia).

A comparison of Eqs. (7) and (8) leads again to the conclusion that

$$P_{max}' \lim_{X\to 0} \bar{P}' = \tfrac{1}{2}P_{max}' \quad (9)$$

provided $\nu_{max}\simeq\nu_o$ and $(\nu_{max}-\nu_o)\ll\nu_{max}, \nu_o$. In order to determine the approximation involved in the use of Eq. (9) for any given case, it is necessary to evaluate and compare the right-hand sides of Eqs. (7) and (8) with $\bar{P}'$ set equal to $\frac{1}{2}P_{max}'$.

The conclusions drawn in the two preceding paragraphs seem to be supported by experimental results obtained for the fundamental vibration-rotation band of CO. For measurements made with low resolution ($\sim$18 cm^{-1}), Eq. (6) was found to be approximately valid in the range of $P_{max}'pl$ between 0 and 1.8‡ for the fundamental of CO in the region between 2018 and 2147 cm^{-1}. Here $\bar{P}'=8.63$ (meters×atmos)$^{-1}$ and $P_{max}'\simeq 18$ (meters×atmos)$^{-1}$ for each of the absorption peaks of the P and R branches, respectively. Detailed calculations for the fundamental vibration-rotation band of CO also indicate that Eq. (8) with $\bar{P}'=\frac{1}{2}P_{max}'$ represents a good approximation to Eq. (7) for values of $P_{max}'X$ up to about 2 or 3.

This analysis clearly indicates that the type of approximation introduced by Schack[9] can be justified readily for apparent absorption coefficients obtained with low spectral resolution. On the other hand, it is obvious that these approximations are of no value whatever in making even semiquantitative estimates of radiant heat transfer§ except at very large pressures where extensive overlapping of rotational lines occurs.[5]

III. AVERAGE ABSORPTION COEFFICIENT EQUAL TO THE RATIO OF INTEGRATED ABSORPTION TO BAND-WIDTH AT ELEVATED PRESSURES

At elevated total pressures, where the ratio of rotational half-width to line spacing is sufficiently small, it can be shown that the apparent transmission τ' is given [8,10] by the approximate relation,

$$2.303\log\tau' = -apl/\Delta\nu, \quad (10)$$

where a is the integrated absorption for the entire vibration-rotation band, $\Delta\nu$ is the effective band width, and τ' is defined by the relation,

$$\tau' = \int_{\Delta\nu} I_\nu d\nu \Big/ \int_{\Delta\nu} I_{o\nu} d\nu \simeq \int_{\Delta\nu} (I_\nu/I_o\Delta\nu)d\nu. \quad (11)$$

‡ The value of P_{max}' was obtained by use of the relation,

$$P_{max}' = \lim_{pl\to 0}\{d\log[1-(I_{\nu max}/I_{o\nu max})]/d(pl)\}.$$

§ This statement is supported, for example, by the fact that data of the type shown in Fig. 1 lead to values for the integrated absorption which are in error by factors of 10 or more. (See reference 8.)

[10] W. M. Eslasser, Harvard Meteorological Studies No. 6 (Milton, Massachusetts, 1942).

Relevant values of τ' for the fundamental of CO at $p_T=700$ psia are summarized¶ in Table II as a function of pl for CO mixed with H_2, He, or A. The value of $\Delta\nu=270$ cm^{-1} was used to calculate τ' from the experimental results. It has been shown[8] that at $p_T=700$ psia the transmission is independent of the nature of the added gas and therefore that all of the experimental data given in Table II can be used in a single plot to test the approximate validity of Eq. (10) and to evaluate a. Thus the values of log τ' for CO−H_2, CO−He, and CO−A mixtures are plotted as a function of pl in Fig. 4.

Reference to Fig. 4 shows that the plot of log τ' *vs* pl is linear over only a relatively small range of optical densities. On the other hand, the limiting slope is seen to pass through the origin and lead to the value $a_F=210$ cm^{-2} atmos^{-1} if $\Delta\nu$ is set equal to 270 cm^{-1}. This value for a_F is seen to be in reasonably good agreement with the more accurate result $a_F=237$ cm^{-2} atmos^{-1}.[8]

TABLE III. Dependence of τ' for the first overtone of CO on pl at $p_T=700$ psia (CO pressurized with He).

pl (cm atmos)	τ'	logτ'
2.71	0.9827	−0.0076
5.43	0.9637	−0.01606
7.01	0.9622	−0.01673
14.0	0.9215	−0.03550
21.0	0.8878	−0.05170
28.0	0.8615	−0.06470
35.1	0.8333	−0.07920

Since the derivation of Eq. (10) involves the assumption that all of the rotational lines of a vibration-rotation band can be considered to be equally intense and equally spaced, the result derived from Fig. 4 must be considered to be satisfactory. Some of the deviations from linearity in Fig. 4 are undoubtedly caused by the fact that the effective band width $\Delta\nu$ is not accurately independent of pl.[1,2] Nevertheless, the data of Fig. 4 supply at least semiquantitative support for the underlying assumption in recently published radiant-heat-transfer calculations, for it appears that, as a first approximation, at large p_T each vibration-rotation band may be represented by a box of width $\Delta\nu$ which is characterized by an average absorption coefficient equal to $a/\Delta\nu$.

Experimental data obtained for the first overtone of CO pressurized with He to $p_T=700$ psia are summarized

¶ All of the experimental data at elevated pressures were obtained by use of the experimental technique described previously. (See reference 8.)

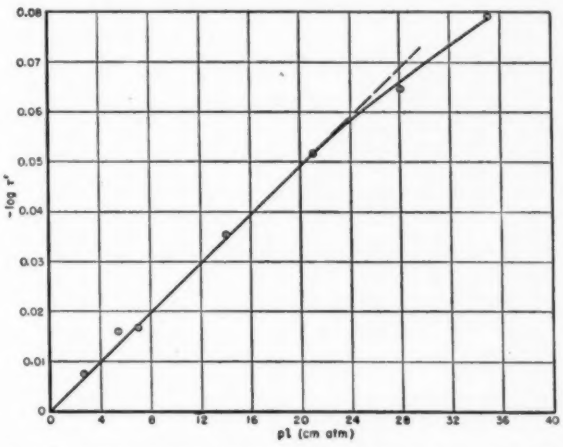


FIG. 5. Plot of logτ' as a function of pl for the first overtone of CO (CO pressurized with He to 700 psia).

in Table III and plotted in Fig. 5. The limiting slope for small pl for the first overtone of CO leads to the conclusion that the integrated absorption $a_o=1.52$ cm^{-2} atmos^{-1}, which is seen to be in good agreement with the more accurate value of 1.64 cm^{-2} atmos^{-1} obtained by other methods.[8] In this connection it may be of interest to note that log τ' is a linear function of pl up to relatively large values for the optical density. Thus it appears that the approximations involved in the use of Eq. (10) hold for a larger range of pl values for the overtone than for the fundamental.

From the preceding results it appears that the approximation $\bar{P}=a/\Delta\nu$ for a given vibration-rotation band will overestimate the actual emissivity by less, the smaller the optical density at a given total pressure. Hence, since the effective band width increases with temperature, it is to be expected that the range of pl values for which $\bar{P}\simeq a/\Delta\nu$ will increase with temperature.

It is apparent that the approximate procedure for radiant-heat-transfer calculations can be improved by replacing the average absorption coefficient by a triangular distribution. In particular, it follows from Eq. (3) that a relation of the form,

$$\tau'\simeq 1-[1-\exp(-P_{\max}'X)]/P_{\max}'X, \qquad (12)$$

would describe the experimental data summarized in Tables II and III. Reasonably good agreement is obtained between values of τ' calculated from Eq. (12) and observed values of τ' if $P_{\max}'$ is set equal to 1.65 cm^{-1} atmos^{-1} for the fundamental. This value of $P_{\max}'$ is obtained by using an equivalent triangular distribution for the observed dependence of the true spectral absorption coefficient P_ν on ν.

The Evaporation Products of Barium Oxide from Various Base Metals and of Strontium Oxide from Platinum*

L. T. ALDRICH†
Department of Physics, University of Missouri, Columbia, Missouri
(Received April 9, 1951)

The evaporation products of BaO heated on ribbon filaments of Pt, Ta, Ni, Mo, and W, and of SrO on Pt have been studied using a high resolution mass spectrometer for positive ion analyses. The ratio Ba^+/BaO^+, was measured for each base material as a function of temperature and of the energy of the ionizing electrons to determine the origin of each ion peak observed. The results indicate: (a) Only on systems BaO on Pt and BaO on Ni, and SrO on Pt were the evaporation processes reproducible without particular attention to procedure. (b) The two systems BaO on Pt and SrO on Pt differ markedly in their evaporation mechanisms. (c) At high temperatures peaks are found due to Ba_2O^{++}, Ba_2O^+, $Ba_2O_2^+$, and other barium compounds depending on the base metals used. (d) A measurable ion current was found due to Sr^+ ions evaporated from Pt.

INTRODUCTION

THE understanding of the physical basis for the electron emission from oxide cathodes has given impetus to all work relating to the evaporation products from BaO and SrO systems. Among the early workers using mass spectrometer techniques to examine such systems were Barton, Harnwell, and Kunsman[1] and Blewett and Jones.[2] More recently Plumlee and Smith[3] have observed the evaporation of the triple oxide (BaSrCa)O from cathodes on various base metals. Grattidge[4] has studied the emission of negative ions from activated cathode systems. Moore and Allison[5] have shown that very thin films of the material evaporated from a SrO-Pt system onto a W or Mo base change the thermionic emission characteristics of these metals in about the same way as does a thick layer of SrO. The following is a report of a mass spectrometric study of the evaporation products using pure single carbonates and some of the base metals which have been used in studying thermionic emission.

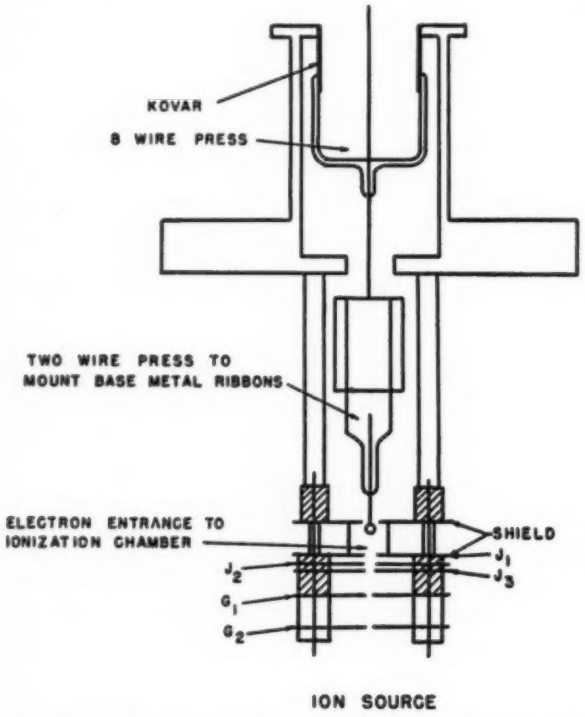


FIG. 1. Ion source for evaporation studies. Spacing between J_1 and J_2, 1 mm. Slit in J_3, 3×14 mm. Slits in G_1 and G_2, 0.25×14 mm. Base metal ribbons, 1.0×0.01 mm, 25 mm long.

EXPERIMENTAL METHODS

A. Instrument

The high resolution mass spectrometer used in this work is similar to that described by Nier.[6] It differs from his only in the source and collector assemblies. The ion source was modified as shown schematically in Fig. 1. The chief alterations in the source were in the slit widths of G_1 and G_2 and in the mounting of the base metal ribbon in the ionization chamber. Our source and collector assemblies were attached to the spectrometer tube using thin aluminum gaskets and were not cooled during baking out periods. A single faraday cage collector with a secondary electron suppressor was mounted behind the slit at the focal point of the spectrometer. This slit was adjustable in width and was set at 0.010 in. for nearly all of the work. A feedback amplifier and linear attenuator were used to measure the collected ion current which permitted accurate current determinations over four orders of magnitude without changing the grid resistor of the preamplifier. The amplified signal was indicated on a Speedomax recorder. A second recording system was used for scanning over a large mass range to permit recording currents of widely varying amplitude. In this case, the attenuator coupled with the recorder varied the fraction of the output signal seen by the recorder from 1 at the low end of the scale to 0.0001 at the high end. This resulted in approximately a loga-

* Supported in part by the ONR.
† Present address: Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, D. C.
[1] Barton, Harnwell, and Kunsman, Phys. Rev. **27**, 739 (1926).
[2] J. P. Blewett and E. J. Jones, Phys. Rev. **50**, 464 (1936).
[3] R. H. Plumlee and L. P. Smith, J. Appl. Phys. **21**, 811 (1950).
[4] W. Grattidge, thesis, University of Manchester (1949).
[5] G. E. Moore and H. W. Allison, Phys. Rev. **79**, 246 (1950).
[6] A. O. Nier, Rev. Sci. Instr. **18**, 398 (1947).

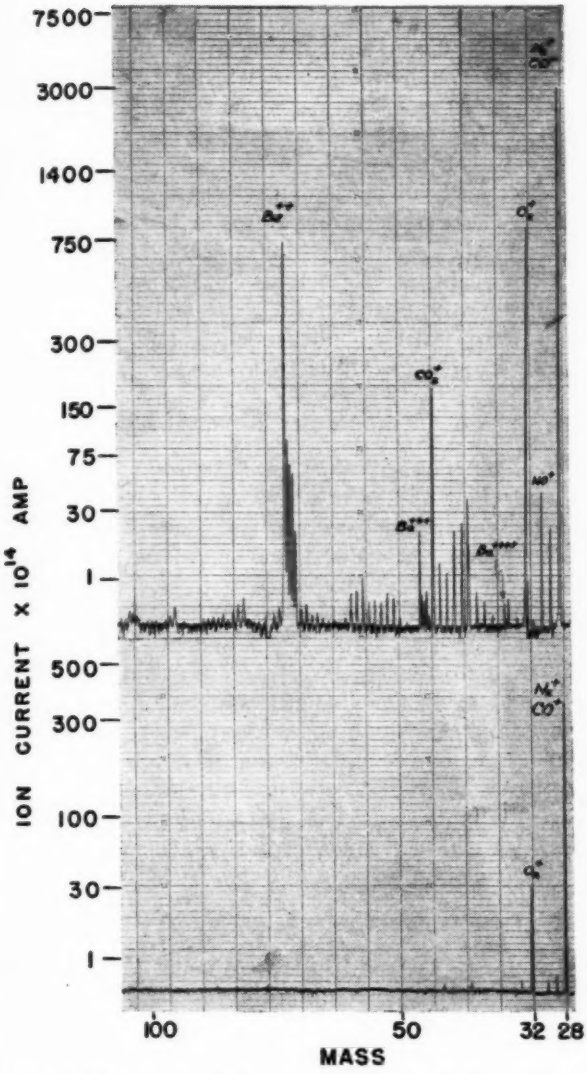


FIG. 2. Mass spectra in low mass region. (a) BaO evaporating from pure platinum at 1400°K, $P=3\times10^{-7}$ mm Hg. (b) Residual peaks, $P=2.5\times10^{-7}$ mm Hg.

rithmic response. The performance of this recording system and the instrument under normal operating conditions is shown in Figs. 2 and 3. To sweep over a given mass region, the accelerating voltage was varied by allowing 9 to 25 microfarads of capacitance (depending on the sweep time desired) to be discharged through a 50-megohm resistor. This sweeping method was used primarily in the search for high mass peaks.

In measuring the Ba^+/BaO^+ ratio, account was taken of the effect of the accelerating voltage on the ion current and the mass 138 and 154 peaks were brought into focus by changing this voltage. No correction due to the effect of the isotopes of oxygen on the BaO^{154} peak was made.

B. Preparation of the System

The following schedule for system preparation was adhered to on each run:

1. Mount base metal ribbon and vacuum fire it in the spectrometer at 1700°K or higher, depending on the metal, for at least fifteen minutes. This was done to supplement chemical cleaning and to remove as much of the alkali metal impurity as possible. Potassium was identified in all of the metals heated in the source including the very pure platinum.‡ Exact heating schedules were worked out by monitoring the potassium peaks while the metal was heated with no coating.

2. Coat the center $\frac{1}{4}$ inch of the ribbon with a distilled water suspension of pure single carbonate.§ This eliminated any residual peaks due to organic binder materials which are normally used in commercial cathode preparation.

3. Assemble the source, mount it in the spectrometer, evacuate and bake the entire system at 250°C for at least four hours with dry ice as the trap refrigerant.

4. After cooling, make test runs on the residual peaks to check the performance of the instrument for resolution and intensity.

5. Put liquid air on the trap and bake again for 30 minutes.

6. Convert the carbonate keeping the pressure below 10^{-5} mm Hg.

6a. One sample of each of the BaO systems thus formed was then heated at 900°K for three hours before proceeding and is denoted as an aged sample in the following discussion.

7. Ba^+ and BaO^+ ion currents were measured for varying temperatures and energies of ionizing electrons. The latter measurements were made at fairly high temperatures so that measurable peaks would be obtained at low (15-volt) electron energies.

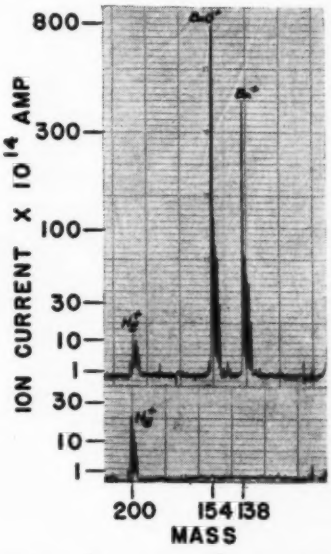


FIG. 3. Mass spectra in high mass region. (a) BaO evaporating from pure platinum at 1500°K, $P=3\times10^{-7}$ mm Hg. (b) Residual peaks, $P=2.5\times10^{-7}$ mm Hg.

‡ Sources of metals used were: Pt—Bell Telephone Laboratories, Inc., Ta, Mo, and W—H. Cross and Company, New York, New York, Ni—Spectrographic Ni—E. M. Wise, International Nickel Company, New York, New York.

§ Mallinckrodt ultra-pure $BaCO_3$. $SrCO_3$ supplied by G. E. Moore, Bell Telephone Laboratories, was of the batch used by Moore and Allison (see reference 5).

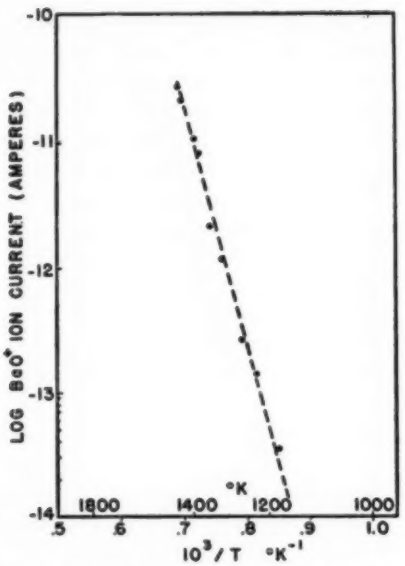


FIG. 4. Log BaO^+ ion current *vs* $1/T$ for BaO on Pt.

8. At high temperatures spectra in the region from mass 154 to 500 were recorded to search for barium compounds. The pressure in the spectrometer rarely exceeded 3×10^{-7} mm Hg after step 6.

EXPERIMENTAL RESULTS

Several processes may take place in the ionization chamber of the mass spectrometer to produce spectra such as those shown in Figs. 2 and 3. Peaks differing from the calculated mass of the material evaporated are often found due to multiple ionization, ion combination, and the combination of ions and neutral molecules. The problem in this work is to determine which ion peaks are characteristic of the evaporation process and the accompanying temperature dependent chemical reactions and which are the result of dissociation and ionization by the electron beam. Any true evaporation process must be temperature dependent and have a characteristic activation energy. Any dissociation produced by the electron beam must depend on the energy of the ionizing electrons according to the ionization efficiency. Unfortunately the latter dependence is difficult to measure quantitatively on small peaks. Data were obtained by varying these two parameters, temperature and electron energy, independently and will be discussed separately for the BaO systems.

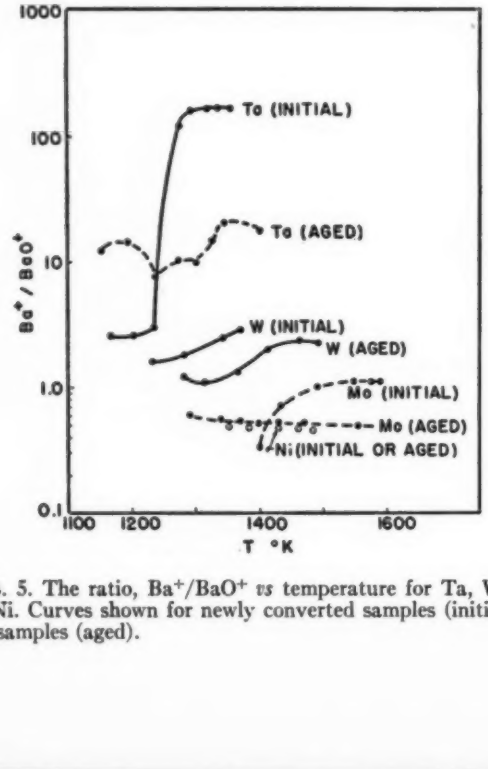


FIG. 5. The ratio, Ba^+/BaO^+ *vs* temperature for Ta, W, Mo, and Ni. Curves shown for newly converted samples (initial) and aged samples (aged).

A. Temperature Dependence

The temperature of the base metal ribbon was varied by controlling the ac heating current through it. The ribbon was connected electrically to the shield of the ion source through the center tap of the ribbon transformer so that the evaporation took place in essentially a field free region. The temperature of the ribbon was measured using a Mo-Ta thermocouple (wires 0.002-in. and 0.003-in. diameter respectively). This thermocouple was calibrated against a Pt-PtRh standard couple up to 1650°K and against an optical pyrometer from 1150 to 2300°K. Figure 4 is a plot of log BaO^+ ion current as a function of the reciprocal of the absolute temperature, for a Pt ribbon coated with BaO. The slope of the line drawn through the experimental points gives the constant K_1, in the equation

$$\log(\text{vapor pressure}) = -(K_1/T)+K_2. \qquad (1)$$

The value of K_1 from the experimental line is 19,000 corresponding to a latent heat of vaporization of 87 kcal/mole and is in good agreement with the value obtained by Claassen and Veenemans.[7] If for this same sample, log Ba^+ ion current is plotted as above, the slope is the same as that for BaO^+ within experimental error. This is one indication that most of the Ba^+ observed results from the dissociation of BaO by electron impact rather than by evaporation.

Figure 5 shows the ratio Ba^+/BaO^+ plotted for the Ta, W, Mo, and Ni base materials and indicates the effect of sample aging. Thus W (initial) indicates that the ratio, Ba^+/BaO^+, was determined directly after the conversion of the carbonate, and W (aged) means that this sample was taken through step 6a in its preparation before these ratios were determined. For the Ta, W, and Mo samples it is seen that the Ba^+/BaO^+ ratio is different for the initial and aged samples. This may be due to the formation, at this relatively low aging temperature, of an interface compound, the presence of which affects the further reaction of the base metal and the oxide. The Ba^+/BaO^+ ratio was not altered by this aging procedure when base metals of pure Ni and Pt were used. The Ba^+/BaO^+ ratio will have its minimum value

[7] A. Claassen and C. F. Veenemans, Z. Physik **80**, 342 (1933).

when the Ba^+ ion arises only from electron dissociation of BaO. This is believed to be the case for the chemically inactive base metals pure Ni and Pt where the Ba^+/BaO^+ ratio is about 0.5 from Figs. 5 and 7. No significance is attached to the point on the Mo (initial) curve of Fig. 5 lying below the ratio 0.5, since the deviation from this value lies within the experimental error at this temperature. The Ba^+/BaO^+ ratio will increase if a chemical reaction occurs between the oxide and base metal producing Ba which evaporates and is ionized along with BaO. This effect is seen to occur to a marked degree on Ta and to a less extent on W and Mo (initial).

The system BaO-Ta, which is of importance as a commercial getter, deviates from the other systems in a marked fashion. Above 1250°K the Ta system in the initial state has an evaporation product which is at least 99 percent Ba. The ratio, Ba^+/BaO^+, changes as the sample is aged and it is not possible to obtain reproducible Ba^+ or BaO^+ ion currents over a temperature cycle for the Ta system in the initial state. An aged Ta system did not show these effects.

BaO-Pt and BaO-Ni systems taken through similar temperature cycles gave identical Ba^+/BaO^+ ratios for all temperatures and the same ion currents whenever the same temperature is reached in the cycle. No positive ion emission from the BaO-Pt system was observed at the highest temperatures at which the system was taken (1500°K).

B. Electron Energy Dependence

Regulation of the ionizing electron current in the source was obtained by controlling the total emission from the filament. This total emission could be kept constant at 300 microamp for a range of electron energies from 12 to 75 volts. However, the geometry of the electron gun was such that the ionizing electron current in the chamber was not predictably related to the total emission, as the energy was changed between these limits. True ionization efficiency curves for a single peak could not, therefore, be obtained by simply varying the electron energy. However, it was possible to compare qualitatively the ionization efficiency of the same peak in different systems, and to obtain quantitative results for ratios of different peaks in the same system. Figure 6 shows a comparison of two such ionization efficiency curves obtained for the ion peak at mass 138, Ba^+, for two of the systems studied. The two curves are normalized to the Ba^+ ion current at 75 volts, since the actual Ba^+ ion current from the Ta system was several times that from the Pt system. Figure 7 is a plot of the log of the ratio, Ba^+/BaO^+, as a function of the electron energy in several systems. Since the ionization energy of Ba is 5.2 ev and the dissociation energy of BaO is 4.7 ev,[8] their sum 9.9 ev represents the energy required to produce Ba^+ from a neutral BaO atom.

[8] A. G. Gaydon, *Dissociation Energies* (John Wiley and Sons, Inc., New York, 1947).

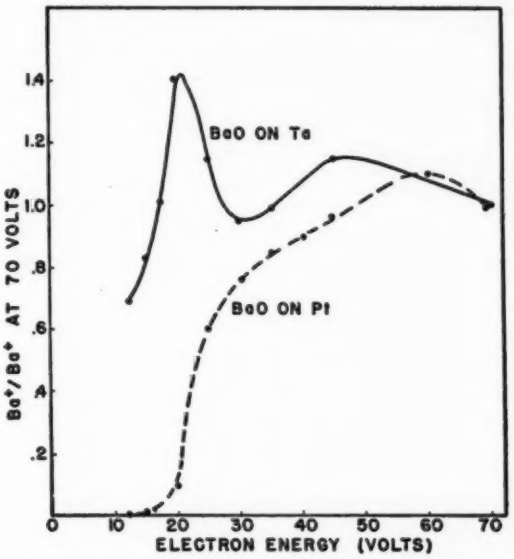


FIG. 6. The ratio, $Ba^+/(Ba^+$ for 70-volt electrons) *vs* electron energy for BaO on Pt and Ta.

Thus at low energy there should be no contribution to the Ba^+ ion current from this dissociation process, and one would therefore expect a decrease in the Ba^+/BaO^+ ratio for low ionizing electron energies. For BaO-Pt, from Fig. 6 it will be seen that the Ba^+ ion current does indeed decrease as this dissociation threshold is approached. Figure 7 shows that the Ba^+/BaO^+ ratio also decreases rapidly in this low energy region, and indicates that less than 1 percent of the evaporated material must come off as Ba. However, for the systems BaO-W, BaO-Mo the Ba^+/BaO^+ ratio does not decrease in the low energy region, signifying that Ba^+ ions are produced

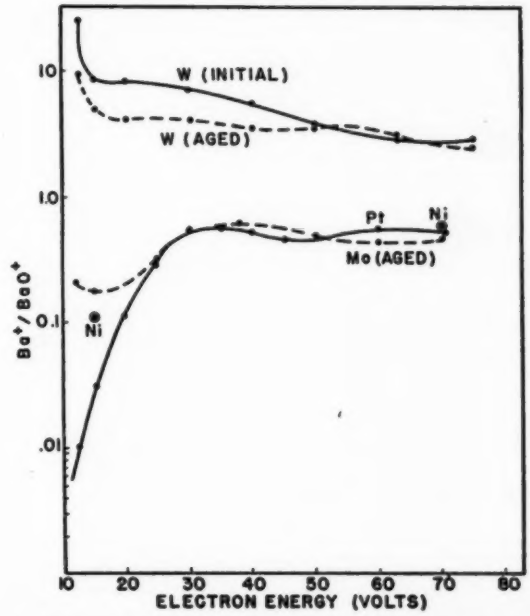


FIG. 7. Ba^+/BaO^+ *vs* electron energy for the systems of BaO on Pt, Mo, W, and Ni (two points only).

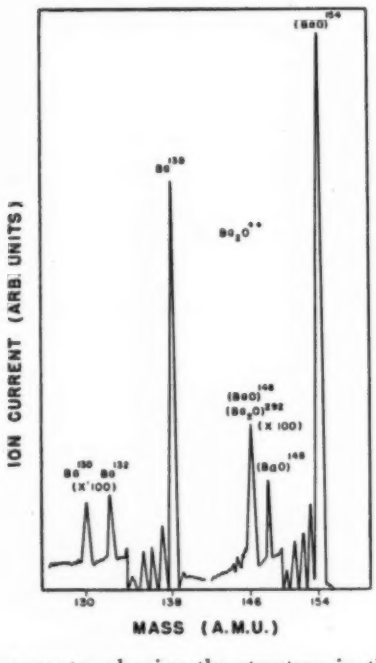


FIG. 8. Mass spectra showing the structure in the region of mass 146.

by a process other than the dissociation of BaO. This undoubtedly represents evaporation of Ba (neutral or ionized) from the system.

C. High Mass Peaks from Various Sources

After the runs had been taken to complete the data above, spectra were taken in the mass range 154 to 500 to see if other ions were present which could be related to a chemical reaction at the interface. While no ion containing Ba and Pt was found from the BaO-Pt system, other high mass peaks were present which could definitely be identified as Ba_2O^{++}, Ba_2O^+, and $Ba_2O_2^+$.

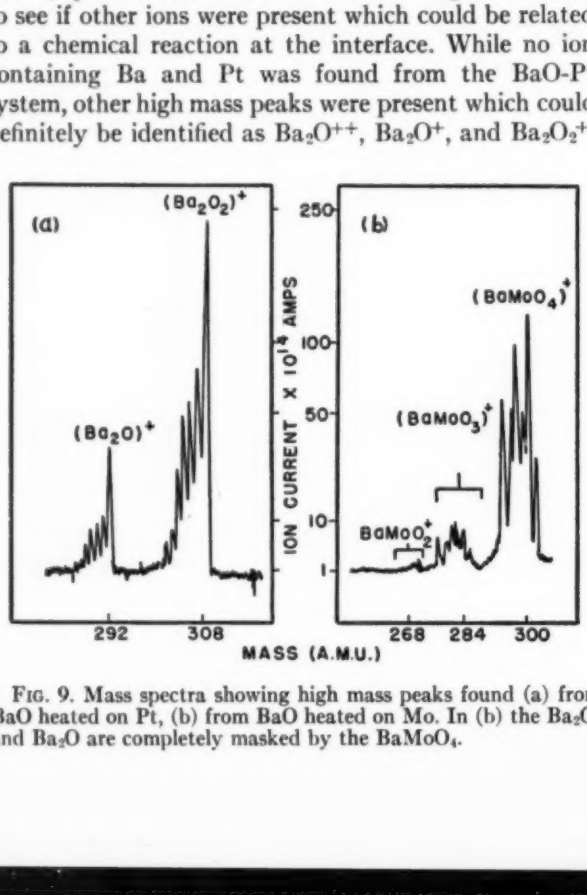


FIG. 9. Mass spectra showing high mass peaks found (a) from BaO heated on Pt, (b) from BaO heated on Mo. In (b) the Ba_2O_2 and Ba_2O are completely masked by the $BaMoO_4$.

Figure 8 shows the peak Ba_2O^{++} (mass 292) which is superimposed on the low mass end of the BaO^+ spectrum at mass 146. For comparison a spectrum of the Ba isotopes is given on the left which shows the Ba^{130} peak as being slightly less than Ba^{132}. A sweep in the mass range above 250 a.m.u. in Fig. 9a shows the presence of Ba_2O^+ and in addition $Ba_2O_2^+$. $Ba_2O_2^{++}$, if it exists, is not differentiated from the large BaO^+ peak at mass 154. The masses present and relative intensities of the peaks in Fig. 9a are in agreement with those predicted by all random combinations of Ba and O isotopes taken in the appropriate proportions. It is possible, of course, that these high mass peaks could be formed by ion-ion combination and ion-neutral atom or molecule combination in the ionization chamber. However, it was found that their intensity had a different temperature dependence than had the BaO^+ ion intensity and therefore could not depend completely on such combination processes for their presence. The ratio of Ba_2O^+ to $Ba_2O_2^+$ was found to be temperature independent, so that Ba_2O^+ is probably an ionization fragment of $Ba_2O_2^+$.

No Ba-Ni compounds were found in two BaO-Ni systems heated to the melting point of the nickel, however, the Ba_2O^+ and $Ba_2O_2^+$ peaks were again present.

On the BaO-Ta systems at temperatures above 2000°K, peaks were positively identified which could be attributed to Ta, TaO, and TaO_2. Other peaks were found just before the Ta ribbon parted and their identification was not as certain, but they correspond in mass to the ions $BaTaO_2^+$, $BaTaO_3^+$ and $Ba_2TaO_4^+$. Since these ions were available only for a short period for observation, their source could not readily be determined and it is possible that some of these are ionization fragments of Ba_2TaO_4.

On the BaO-Mo system ions of $BaMoO_4^+$ were found above 1600°K and identified by the number and relative heights of the peaks which are the result of the combination of the five main Ba isotopes and the most abundant Mo isotopes. Spectra in the mass region 260 to 315 are shown in Fig. 9b. The sub-oxides which are apparent in this figure are probably due to the electron dissociation of $BaMoO_4$. Even at the highest temperatures no peaks were found in the mass regions corresponding to Mo^+, or any oxides of Mo. This differs from the behavior of both Ta and W in their evaporation of the metal and metal oxides and indicates the great stability of the molybdate.

On the BaO-W system a Ba-W compound was found at temperatures above 1600°K which corresponded in mass to $BaWO_4$. This peak at mass 380 was just beyond the limit of resolution of the spectrometer so that isotopic identification was not possible. Other ion peaks were noted which could be attributed to $BaWO_3$, $BaWO_2$, BaWO, W, and WO_2. WO if present would have occurred among the residual Hg peaks and if it existed, it was not large enough to be distinguished above this background.

Solid-state chemical reactions between alkaline earth oxides, and a variety of base metals have been studied by Rooksby. He finds in the cases of W and Mo that the compounds[9] formed are probably Ba_3WO_6 and Ba_3MoO_6. On a base metal of Ta, Affleck[10] finds that one or more compounds are formed but their identification is not certain. The compounds reported by Rooksby are just above the upper limit of the spectrometer (500 amu).

D. SrO on Pt

This system is sufficiently different from any of the BaO systems to merit separate presentation. The temperature dependence of Sr^+ and SrO^+ ion currents is difficult to observe due to the relatively low vapor pressure of SrO for temperatures below the melting point of Pt. To obtain a temperature calibration on this system‖ a blank run was made with the ionization chamber mounted in an evacuated sealed-off glass tube equipped with a flat glass window. The underside of the Pt strip was coated with SrO. Optical pyrometer measurements of the ribbon temperature were taken as a function of the heating current through the ribbon. After these data were obtained, a Pt/Pt-10 percent Rh thermocouple was attached to the ribbon and the optical pyrometer readings were repeated and correlated with thermocouple data. From these two sets of observations one establishes the relation between the ribbon current and its temperature under the standard operating condition, i.e., without thermocouple attached.

Whereas in the BaO-Pt system the BaO^+ ion current exceeded the observed Ba^+ ion current, in the SrO-Pt system the Sr^+ ion current was always an order of magnitude greater than the SrO^+ ion current. Over the temperature range 1725–1810°K the slope of a plot of log Sr^+ ion current *vs* $1/T$ is 30,000 corresponding to a heat of vaporization of 144 kcal/mole. A plot of log SrO^+ ion current *vs* $1/T$ had a similar slope within the limit of experimental error (10 percent). However, this observation does not necessarily imply that Sr^+ ions are derived from the electron dissociation of SrO since the Sr^+/SrO^+ ratio changes from 34 at 1725°K to 21 at 1810°K.

When attempts were made to measure the ratio Sr^+/SrO^+ *vs* electron energy, as in the case of Ba, the ratio increased, with decreasing electron energy. These data were difficult to obtain because the SrO^+ peak was always small; however, the increasing ratio with decreasing electron energy suggested that a residual Sr^+ ion current might be observed for no ionizing electron current. Such a residual current was found amounting to about 10 percent of the total Sr^+ ion current obtained with 1.5 ma ionizing electron current. This ion current had essentially the same temperature dependence as that found for Sr^+ ions formed by electron bombardment discussed previously.

From these data it seems apparent that a much different situation exists than was found in the BaO-Pt system. One possible explanation which is consistent with both the observed high Sr^+/SrO^+ ratio and the temperature dependence of this ratio is that thermal dissociation of SrO occurs in the heated sample according to

$$SrO \rightarrow Sr + \tfrac{1}{2}O_2. \qquad (2)$$

The observed thermionic emission of Sr^+ ions even though they constitute an insignificant fraction of the total Sr evaporation (<0.01 percent) would seem to indicate that reaction (2) is energetically possible. This reaction predicts that in addition to the already observed evaporation products SrO, Sr, and Sr^+, oxygen should also be detected. It was observed that there was indeed an increase over background of the peaks $[(O^{16})^+ + (O^{16})_2^{++}]$ and $(O^{16})_2^+$ during the evaporation process. The sum of these two ion currents was about one-half of the observed Sr^+ ion current. Since the $(O^{16})^-$ peak could not be observed and since the ionization efficiency of O is probably less than that of Sr, these data seem to be in agreement with the postulated reaction (2). The close agreement of the measured heat of vaporization of Sr 144, kcal/mole, and the reported value[7,11] of 141 kcal/mole for SrO suggests that the same evaporation product was involved in each case. In the experiments of Claassen and Veenemans, and Moore, *et al.* the evaporation took place in sealed-off tubes such that Sr evaporated from the heated sample to the cold collector walls could readily recombine with the O released. Thus even though the beginning and end product is SrO, the evaporation products may be Sr and O as found in this study.

SUMMARY

The products evaporated when BaO is heated have been shown to depend on the particular base metal on which it is heated and on the temperature. In the case of Pt at least 99 percent of the evaporated material is BaO, whereas in the case of Ta at least 99 percent is free Ba. Chemical compounds of Ba with oxygen and with the base metal can be evaporated at high temperatures. The compound in the case of Mo is easiest to evaporate and the Ta compound (or compounds) the most difficult. Since these compounds are evaporated only at temperatures much higher than those of normally operating cathodes, their presence as interface compounds can only be inferred and these observations are probably only valid when they confirm results of nondestructive measurements.

The results found with the SrO-Pt system indicate the thermal evaporation of Sr^+ ions and the existence of a high Sr^+/SrO^+ ratio. This result is in marked contrast to the BaO-Pt system.

[9] H. P. Rooksby and E. G. Stewart, Nature **157**, 548 (1946).
[10] J. H. Affleck (private communication).
‖ I am indebted to Dr. Walter Grattidge for this calibration.
[11] Moore, Allison, and Struthers, J. Chem. Phys. **18**. 1572 (1950).

ACKNOWLEDGMENTS

It is a pleasure to acknowledge financial support for the construction of the mass spectrometer from the University of Missouri Research Council and the Research Corporation. The mechanical assembly of the critical parts of the spectrometer tube were chiefly the work of R. B. Thorness, E. F. Cave, and O. Roloff. C. E. Smith helped with electronics assembly and in the early stages of getting the instrument in operation. Finally, special thanks are due to my colleague, A. S. Eisenstein, for his continued support and helpful criticism of this work.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 9 SEPTEMBER, 1951

The Allotropic Transformation of Hafnium

POL DUWEZ*
California Institute of Technology, Pasadena, California†
(Received April 12, 1951)

The existence of an allotropic transformation in hafnium, suggested by Zwikker in 1926, has been confirmed. The transformation temperature is 1310±10°C. The high temperature beta-form is probably body-centered cubic.

INTRODUCTION

IN a paper published in 1926, Zwikker[1] reported for the first time the existence of an allotropic transformation in zirconium and in hafnium. His conclusions were based on anomalies found in the variation of electrical resistivity with temperature. In the case of zirconium, Zwikker's measurements indicated a maximum around 877°C followed by a minimum at 1157°C and he concluded that a transformation was taking place between these temperatures. It has been shown since[2] that the allotropic transformation in zirconium takes place with little or no hysteresis at 862±5°C. From the maximum and minimum observed in the hafnium electrical resistance curves, Zwikker concluded that a transformation occurred between 1327 and 1527°C. No further reference to work on the transformation of hafnium has been found in the literature.

The results of the present investigation prove that an allotropic transformation takes place in hafnium at 1310±10°C. In addition, a study of the hafnium-columbium system led to the conclusion that these two metals form a complete series of solid solutions at high temperature. This result is interpreted as an indirect proof that the high temperature form of hafnium is body-centered cubic.

THERMAL ANALYSIS MEASUREMENTS

The hafnium metal used for this investigation was prepared by the Van Arkel process by the Foote Mineral Company (bar No. 778). The lattice parameters were measured from a powder pattern obtained with a 14.32 cm camera and copper $K\alpha_1$ radiation. The Cohen least-squares method was applied to lines (123), (302), (006), (205), and (220) in the back reflection range. These parameters are $a=3.1952$A, $c=5.0569$A, and $c/a=1.5827$.

It was inferred that if a transformation existed in hafnium, it would be of the same type as that found in titanium and zirconium. The transformation of the hexagonal close-packed to the body-centered cubic high temperature form of these two metals is diffusionless and practically independent of the rate of cooling.[3] The technique used for the study of titanium and zirconium was therefore applied to hafnium.

This technique, described by Greninger in a study of the martensite transformation in steels,[4] consists of heating a very small specimen of metal to which fine thermocouples are spot welded and quenching it by means of a blast of helium. From the temperature *vs* time curve recorded on an oscillograph, the transformation temperature and the rate of cooling may be determined. The results of measurements at various rates of heating and cooling are given in Table I. It appears that the transformation temperature is not affected by the rate of cooling. The transformation is therefore of the diffusionless type similar to that found in titanium and zirconium.[3]

CRYSTAL STRUCTURE OF THE HIGH TEMPERATURE FORM OF HAFNIUM

Because of the difficulties involved in obtaining diffraction patterns at high temperature without contamination of the specimen, an indirect method was used to establish the crystal structure of beta-hafnium.

* Associate Professor of Mechanical Engineering.

† This work was part of a research program on the allotropic transformation of titanium sponsored by the ONR, U. S. Navy, under contract number N6onr-24430. The pure hafnium was obtained through the courtesy of Colonel B. R. Lawrence, Jr., Chief, Materials Laboratory, Engineering Division, Wright-Patterson Air Force Base, Dayton, Ohio. The author wishes to thank Mr. William V. Wright, Jr., for helping in carrying out the experiment.

[1] C. Zwikker, Physica **6**, 361–365 (1926).
[2] W. G. Burgers, Physica **1**, 561 (1934).
[3] P. Duwez (to be published).
[4] A. B. Grenninger, Trans. Am. Soc. Metals **30**, 1 (1942).

TABLE I. Effect of rate of cooling on the allotropic transformation of hafnium.

Rate (°C/sec)	Temperature (°C)
Heating	
6	1305
7.5	1323
8	1315
Cooling	
10	1310
16	1310
20	1300
20	1300
20	1305
21	1323
22	1314
25	1323
40	1310
46	1300
80	1325
280	1304
330	1306
1250	1300
5200	1310
6100	1315
7100	1308
8200	1315

On the basis of the Hume-Rothery rules governing alloy formation, it can be predicted that if beta-hafnium is body-centered cubic, alloys of hafnium with either columbium or tantalum should form a continuous series of solid solutions at high temperature. A series of hafnium-columbium alloys were prepared by melting in a helium arc furnace similar to that described in reference 5.[5] Since the total amount of hafnium available was less than 5 grams, the weight of each alloy sample was limited to approximately 0.5 gram. After melting, the alloys were sealed in evacuated fused silica tubes, homogenized for 48 hours at 1000°C, and quenched from that temperature.

Nine compositions were investigated: 10, 20, 23.5, 30, 43.5, 47, 55, 66, and 77 atomic percent columbium in hafnium. The x-ray diffraction patterns of the alloys containing more than 30 percent columbium were readily indexed on the basis of a body-centered cubic structure. The patterns of the three alloys having less than 30 atomic percent columbium contained reflections of hexagonal hafnium, in addition to the reflections of a body-centered cube. The plot of lattice parameter of the body-centered cubic phase *vs* composition is shown

[5] Schramm, Gordon, and Kaufmann, Trans. Am. Inst. Mining Met. Engrs. **188**, 195 (1950).

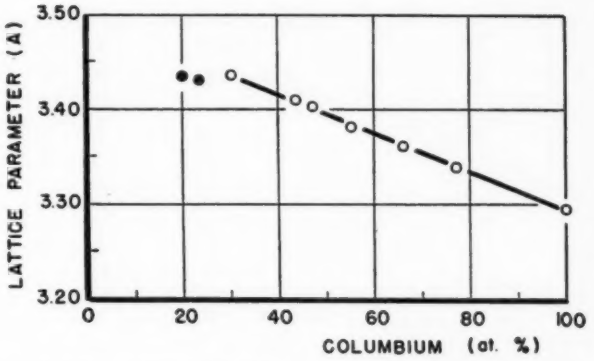


FIG. 1. Lattice parameter of hafnium-columbium alloys. (○) one-phase alloys (BCC). (●) two-phase alloys (BCC+HCP).

in Fig. 1. The diagram clearly demonstrates that at 1000°C a solid solution exists between hafnium and columbium from 30 to 100 percent columbium, and a two-phase region (hexagonal hafnium plus cubic solid solution) extends from an undetermined concentration up to about 30 percent columbium.

The extensive solid solubility of hafnium in columbium at 1000°C suggests that at temperatures above the transformation of hafnium the two metals are completely soluble in each other. This statement, however, cannot be checked by room temperature x-ray diffraction measurements because the high temperature body-centered cubic phase in the alloys rich in hafnium cannot be retained by quenching. It was observed that when the alloys containing 10 and 20 atomic percent columbium were rapidly cooled from the melt, the resulting structure was a mixture of hexagonal close-packed and body-centered cubic phases. This behavior is exactly the same as that observed in the titanium-molybdenum system,[3] in which there is complete solubility at high temperature, but the body-centered cubic structure cannot be retained by quenching unless the molybdenum concentration is greater than about 6 atomic percent. For concentrations smaller than 6 percent, the transformation into an hexagonal supersaturated solid solution always takes place, at least partially, through a martensitic type of transformation. By extrapolating the curve of Fig. 1 to zero columbium concentration, the lattice parameter obtained is 3.50A, which represents approximately the edge of the unit cell of the body-centered cubic modification of hafnium if it would be measured at room temperature.

Electrical Charge Storage in Polystyrene Capacitors

L. A. MATHESON AND V. J. CALDECOURT
The Dow Chemical Company, Midland, Michigan

(Received May 11, 1951)

Measurements on charge storage in polystyrene capacitors over a period of months indicate that exceptionally fine capacitors may be made of polystyrene film when sufficient care is taken. Very small charge "soakage" occurs in such a capacitor, and charges may be retained for times of the order of one hundred years and possibly longer. The specific resistance approaches 10^{22} ohm cm when measured over a period of months. The power factor of polystyrene is about 10^{-4} for frequencies of 10^{-1} to 10^{-6} cycle per sec with a tendency to increase at the lowest frequencies and with a possible peak at still lower frequencies. The difference between charge soakage during charging and the charge recovery after shorting indicates some nonreversible current flow when measurements are made over several months' time.

ELECTRICAL capacitors made with polystyrene foil as the dielectric have been recommended highly for charge storage purposes and have been used in highly critical locations such as gun directors and computing machines.[1,2] The low dielectric absorption or polarization of polystyrene results in a small charge "soakage" and later release. A capacitor seems ideally suited for dielectric measurements, since it possesses low ratio of surface leakage to volume conductance and has sufficient capacitance that available equipment may be used for measurement. Figure 1* compares a polystyrene capacitor with capacitors using several other dielectrics. The condensers were charged 30 seconds, shorted for the times noted on the abscissas, and the charge collected in the next 30 seconds plotted as ordinate. Polystyrene has very high insulation resistance and shows very little current soakage, less than 1/10 that of competitive materials tested.[2]

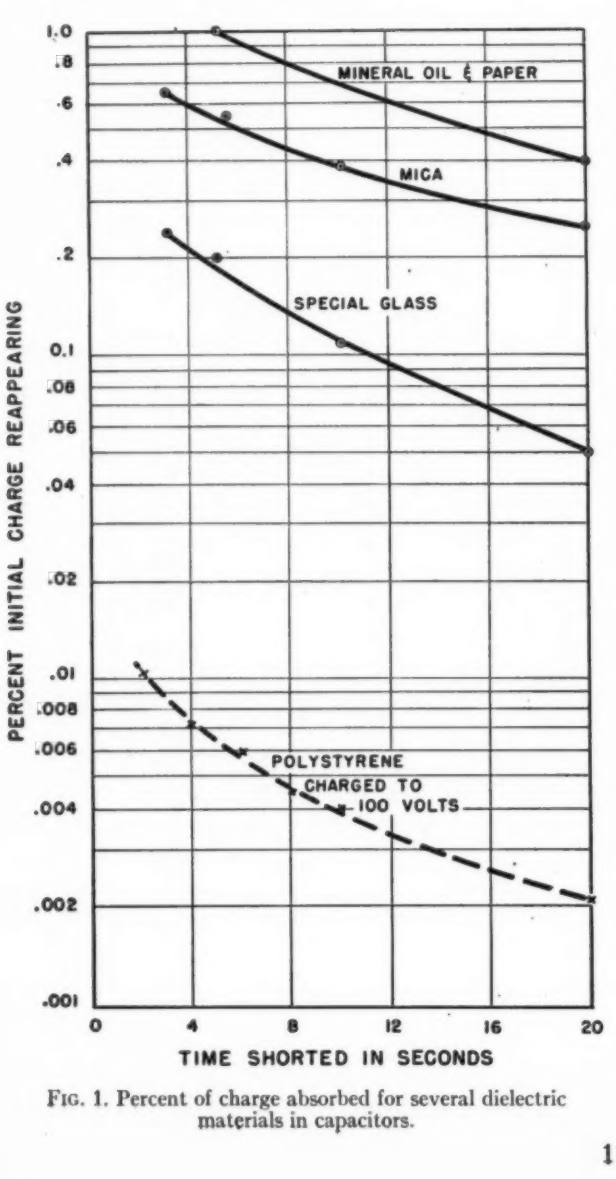


FIG. 1. Percent of charge absorbed for several dielectric materials in capacitors.

We have made measurements on several polystyrene capacitors over longer periods of time at room temperature, about 25°C. They were charged to 135 volts and observed while disconnected for one month. One capacitor (obtained from John E. Fast and Company, Crawford Avenue, Chicago, Illinois) showed exceptionally small loss of voltage. No provable voltage drop occurred during the first month. Upon careful measurement it was observed to lose 0.45 volt in the following 75 days. This performance means a time constant of about 60 years or, if the voltage drop continued at the same rate, that about 40 years would be required for the capacitor to discharge to half the initial voltage. The insulation resistance is greater than 10^{15} ohms/microfarad for this long observation time. From this and the dielectric constant of 2.5 one can calculate an average specific resistivity of 9×10^{21} ohm cm for the polystyrene dielectric. This charge and voltage retentivity should find application for a variety of purposes such as charge or voltage standardization.

To elucidate the mechanism of the loss of charge, absorbed charge was measured as a function of the time the capacitor was held charged, see Fig. 2. A capacitor was charged for a time long with respect to measure-

[1] Saunderson, Caldecourt, and Peterson, J. Opt. Soc. Am. 35, 681 (1945).

* From a figure in a forthcoming American Chemical Society Monograph *Styrene* to be published by the Reinhold Publishing Corporation, New York.

[2] J. R. Weeks, Bell Labs. Record 24, 111 (1946).

ment times, discharged, and measurements of voltage reappearing plotted on a log time scale. It is seen that straight lines result. The time of unshorting was not accurately controlled, so the times of zero voltage vary and the heights of the lines are not significant. All data on Fig. 2 were obtained with the same capacitor.

These discharge curves, according to the principle of superposition should reproduce charge soakage phenomena during charging.[3] The voltage fall during holding of a charged condenser is caused by this charge soakage plus nonreversible phenomena such as ionic conduction. The reversible part of the charge, i.e., that which will reappear after shorting would appear to be not more than 50 mv per half-decade of the time scale of Fig. 2. Therefore, the 0.45-volt loss of the capacitor occurring from the 30th to 105th day after charging is apparently mostly irreversible, for instance, ionic conduction.

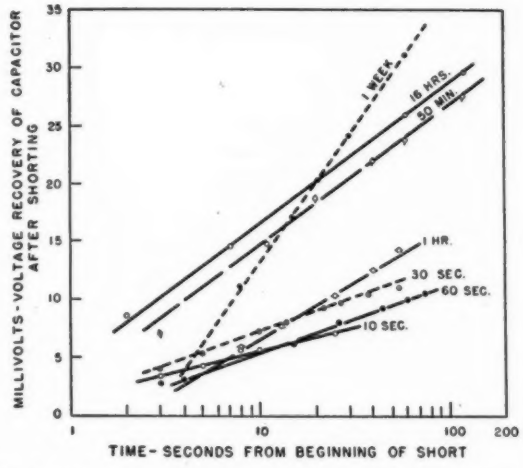


FIG. 2. Voltage recovery of polystyrene capacitor. Charged to 135 volts for times indicated on lines. Shorted at time=0 and unshorted for observation of charge release.

The reversible part of the charge as studied is directly related to the change of dielectric[4] constant, ϵ^1 with increase in time, t. It is possible to obtain an approximate value of the power factor, $\tan\delta$, from the slopes of the lines in Fig. 2, i.e., $d\epsilon^1/d\log t$, and the approximate formula

$$\tan\delta=\frac{\pi}{2\epsilon'}\frac{d\epsilon^1}{d\ln t}=0.68\frac{d\epsilon'}{d(\log t)}. \qquad (1)$$

The derivation of (1) assumes constancy of $d\epsilon^1/d\log t$ over a range of t. The various slopes in Fig. 2 are plotted *versus* log of time of charge in Fig. 3. Formula (1) was used to calculate an approximate power factor as shown on the right-hand scale of Fig. 3.

[3] M. F. Manning and M. E. Bell, Revs. Modern Phys. **12**, 215 (1940).

[4] C. G. Garton, Trans. Faraday Soc. **42A**, 56 (1946).

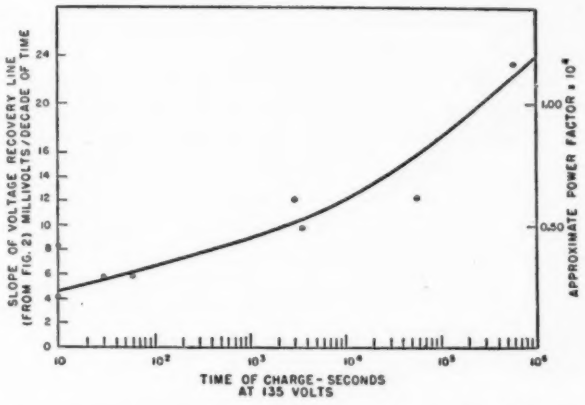

SLOPE OF VOLTAGE RECOVERY LINE (FROM FIG. 2) MILLIVOLTS/DECADE OF TIME
APPROXIMATE POWER FACTOR × 10⁴
TIME OF CHARGE - SECONDS AT 135 VOLTS

FIG. 3. Slopes of lines in Fig. 2 plotted *vs* log of time of charge. Power factor is calculated by formula (1).

This power factor of about 10^{-4} is approximately the same as that measured from 10^2 to 10^9 cycles.[5] Apparently this magnitude of power factor extends even on into the microcycles. The cause of this background power factor is still uncertain. (1) Purification of the dielectric has decreased the power factor and might do so still further. To obtain the results quoted above, great care is required in purification and preparation of monomer and polymer and during fabrication of the capacitors. It might be noted again that the reason for the variations in slope in Figs. 2 and 3 is unknown. On the basis of simple dielectric theory one would expect the effect of longer charging time to be mainly an activation of the slower degrees of freedom which should be evidenced by longer time of recovery rather than greater initial rate of recovery.

It is noted that the slopes in Fig. 2 increase with increased charging time as if the effective power factor increased slowly with increasing time of polarization. Figure 3 shows a plot of slope in Fig. 2 *versus* log of time of polarization. The curve rises sharply at longer times. This seems contrary to a picture of polarization with a relatively uniform frequency distribution which would predict a simple straight line in Fig. 3. It may be that a

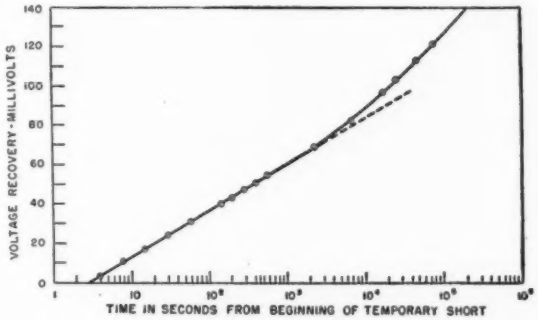


FIG. 4. Voltage recovery on polystyrene capacitor charged 1 week at 135 volts and shorted for several seconds starting at time=0.

[5] A. von Hippel and L. G. Wesson, Ind. Eng. Chem. **38**, 1121 (1946).

power factor maximum is being approached with its peak at less than 10^{-6} cycle.

Measurements of the absorbed charge were made on the capacitor mentioned above which had been charged for 1 week. The quantity of electricity released from the discharged capacitor is shown in Fig. 4 *vs* log time. The plot is not a simple straight line as observed in Fig. 2. The slope increases with time, indicating an effectively higher power factor at the lower frequencies, of the order of 1 microcycle or 10^{-6} cycle. This increased power factor may be related to the factor causing the increased slope in Fig. 3 as the time of charge is increased.

SUMMARY

The measurements indicate that exceptionally fine capacitors may be made of polystyrene film when sufficient care is taken. Very small charge "soakage" occurs in such a capacitor and charges may be retained for times of the order of one hundred years and possibly longer. The specific resistance approaches 10^{22} ohm cm when measured over a period of months. The power factor of polystyrene is about 10^{-4} for frequencies of 10^{-1} to 10^{-6} cycle per sec with a tendency to increase at the lowest frequencies and with a possible peak at still lower frequencies.

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 9 SEPTEMBER, 1951

Some Experimental Indications of the Stresses Produced in a Body by an Exploding Charge*

JOHN S. RINEHART
Michelson Laboratory, U. S. Naval Ordnance Test Station, Inyokern, China Lake, California
(Received April 4, 1951)

The effects produced by small cylindrical charges that were detonated on the surfaces of heavy steel plates are discussed. Particular attention is paid to (a) shapes of crater, (b) changes in hardness, (c) flow patterns, (d) fractures, and (e) changes in microstructure. The distribution of stress appears in some respects to correspond to that which might be set up by a static load. The mechanisms of the inelastic deformation, however, differ very greatly between the static case and the dynamic case. Dynamic loading produces numerous shock twins and fractures that undoubtedly occur because of the very high strain rates involved.

INTRODUCTION

WHEN an explosive charge is detonated on the surface of a body, a large amount of energy is introduced into the body in the form of a high intensity compressional stress wave of short duration.[1] The energy in this wave distributes itself within the body and eventually becomes dissipated. The absorbed energy appears, at least in part, in the form of observable inelastic deformations. A careful examination of these deformations throws considerable light on the probable nature of the stress distribution that must have existed in the body after detonation of the charge. The effects produced by small cylindrical charges that were detonated on the surfaces of large annealed medium steel plates will be described in this paper.

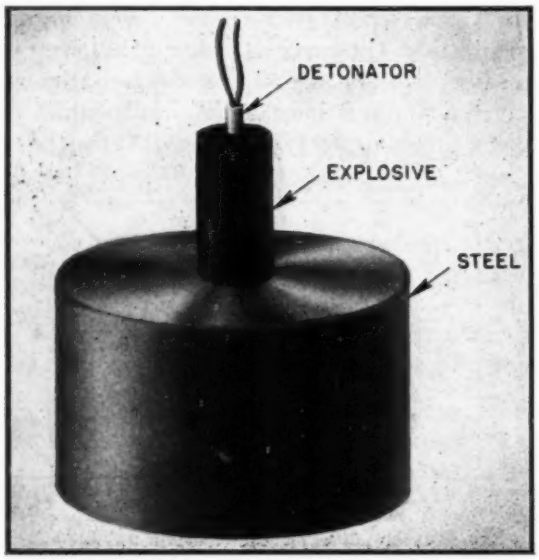


FIG. 1. Experimental arrangement for applying charge to steel plate.

The experimental arrangement is shown in Fig. 1. The explosive was Composition C3. Two-inch long cylindrical charges of four different diameters, ½ in., 1 in., 1½ in., and 2 in., were fired. The steel plates were 2 in. thick and 5 in. in diameter. Each plate was sectioned after the explosive had been detonated.

THE CRATERS

The most obvious effect of the explosion is the production of a depression or crater whose area corresponds roughly to the area of the base of the explosive cylinder. Sectional views of four typical craters are reproduced in Fig. 2. The shape of the crater is seen to

* Work supported by Armament Branch, ONR.

[1] The pressures produced by exploding charges are exceedingly high, probably of the order of 4,000,000 lb/in.^{2} On the other hand, they drop to zero within a few microseconds.

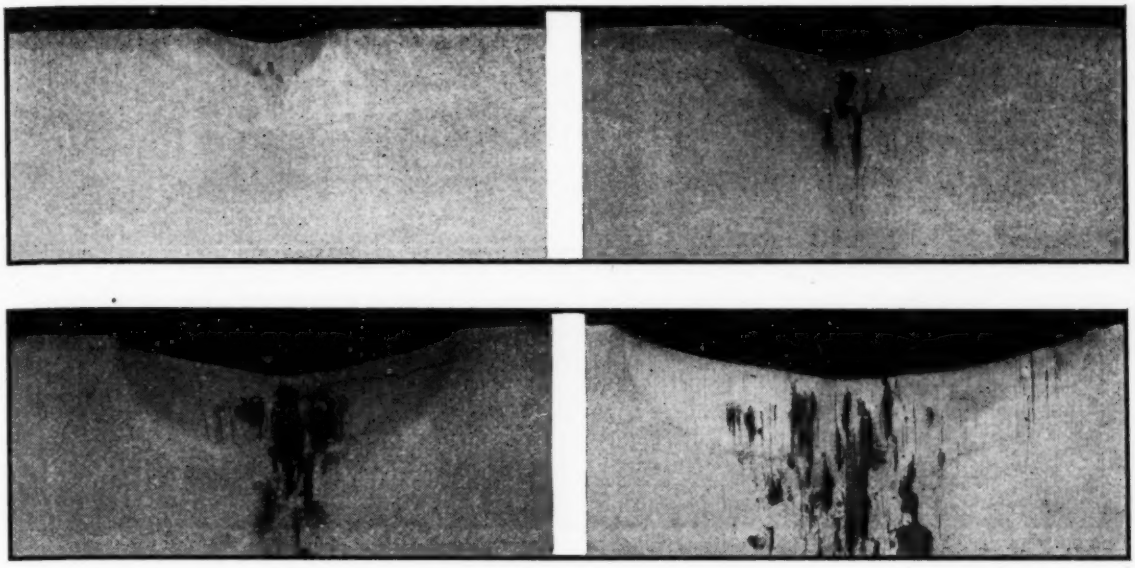


FIG. 2. Sectional views of four typical craters. Upper left: ½-in. diameter charge. Upper right: 1-in. diameter charge. Lower left: 1½-in. diameter charge. Lower right: 2-in. diameter charge.

be substantially the same in each case. It is not the type of crater that a high velocity missile produces when it strikes a steel plate or that a rock makes when thrown into mud. Rather, the crater is nearly conical with almost straight sides. The depth of the crater and the cone angle both increase with increasing charge diameter. Typical values for these quantities are listed in Table I.

TABLE I. Summary of measurements made on craters and highly worked region.

Diameter of charge (in.)	Depth of crater (in.)	Total cone angle of crater (deg)	Thickness of highly worked region (in.)	Volume of crater ($in.^3$)	Volume of highly worked region ($in.^3$)
0.5	0.059	152.0	0.157	0.002	0.032
1.0	0.138	156.0	0.258	0.073	0.231
1.5	0.216	159.0	0.334	0.162	0.784
2.0	0.276	161.5	0.394	0.510	1.67

It seems reasonable to suppose that the shape of the crater reflects the pressure distribution that must have existed on the surface of the steel plate. It is not possible to say with certainty, however, whether it is the magnitude or the duration of the pressure that is of the most significance here, since so little is known about the way in which steel behaves under such rapid and intense loading. Very likely, the depth of the crater at a particular point is proportional to the total impulse, i.e., average pressure times time, that acted at that point.

The way in which the metal moves to form the crater is apparent in Fig. 3. One-inch diameter charges were fired on two plates. The rolling texture of one plate was perpendicular to the direction of application of pressure and of the other, parallel to the direction of the pressure. The plates were then etched to bring out the texture. Generally, the metal is seen to have moved in the direction of application of the pressure. Appreciable lateral motion occurred only near the edges of the crater. There appears to be no splashing of the metal.

FRACTURES

The over-all effect of the explosive is illustrated in Fig. 4. The four sections are typical of plates acted

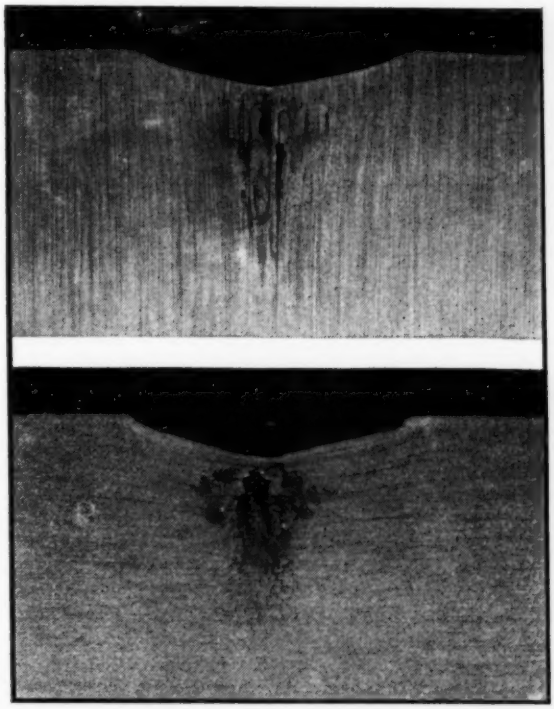


FIG. 3. Sectional views of two plates acted upon by 1-in. diameter charges. Upper: rolling texture parallel to applied pressure. Lower: rolling texture perpendicular to applied pressure.

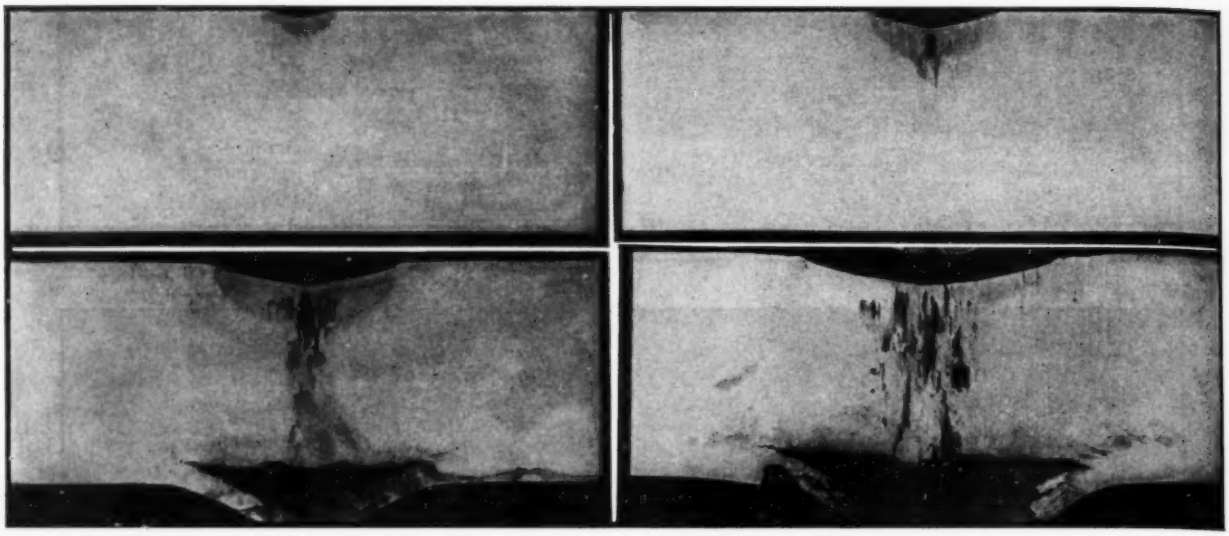


FIG. 4. Sectional views that show over-all effect of explosives. Layout of plates same as in Fig. 2.

upon by ½-in., 1-in., 1½-in., and 2-in. diameter charges. The shapes of the craters have already been discussed. Scabbing, or spalling, is seen to occur in two of the cases. This type of fracture occurs as a result of reflection of the high intensity compressional stress wave at a free surface. Factors upon which scabbing depend have been studied extensively and will be the subject of a later paper.

Many fractures, parallel to the direction of the applied stress, are generated within the plate. The size, number, and extent of these are seen to increase with increasing charge diameter. They are brittle-type fractures. The tensile stresses that produce them probably originate in the dilation that accompanies application of a compressive stress. The tensile stresses will be expected to act at right angles to the direction of the applied compressive stress. Under slow loading rates the material would flow plastically in a lateral direction. At very high loading rates steel is known[2] to behave more nearly as a brittle than as a ductile material; hence, fracturing and not plastic flow might be expected to occur.

The internal fractures do not extend to the surface of the plate on which the explosive was placed. The very high pressure produced by the explosive seems to prevent in some manner the formation of fractures in this region. It may be that initially the material fractures and then heals by flowing plastically. The exact nature of the stress condition existing near the surface is not well defined. It is not unconceivable that Bridg-

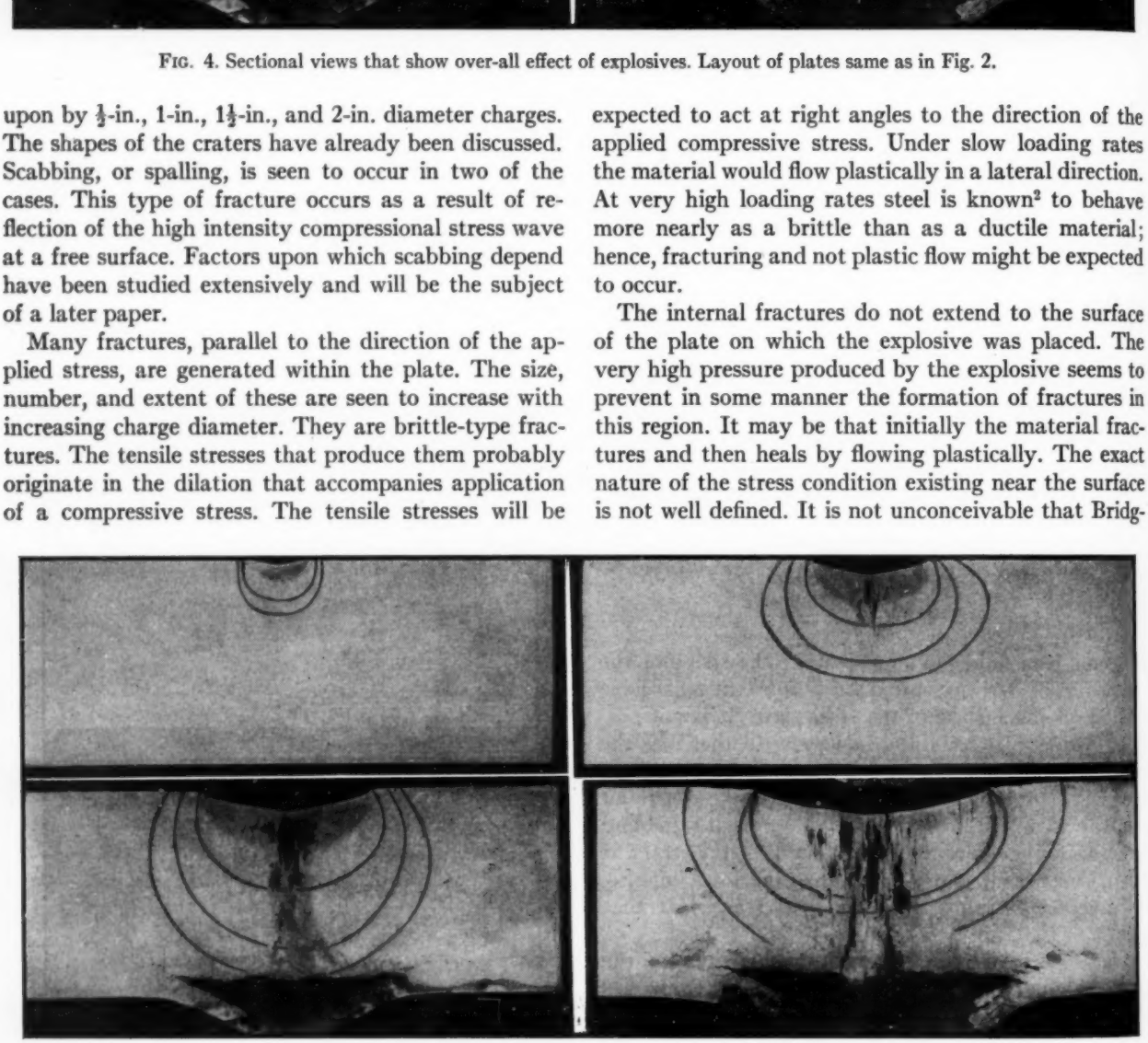


FIG. 5. Sectional views on which contours of equal hardness have been traced. Layout of plates same as in Fig. 2.

[2] D. S. Clark and D. S. Wood, Trans. Am. Soc. Metals **42**, 45 (1950).

man's observation[3] that tensile fractures occur with difficulty under very high hydrostatic pressures is related to the present observation although no definitive statement can be made in this regard.

PLASTIC DEFORMATIONS

Plastic deformation also takes place within the plate. Of particular interest is a clearly defined region, evident in the photographs of Fig. 4, that surrounds the crater and, when etched heavily, contrasts markedly with the remainder of the section. The volume of this region is roughly four to five times that of the corresponding crater. The metal within this region has undoubtedly undergone severe cold-working. The surface that bounds the region intersects the plate of the section along a curve that approximates an ellipse. A reasonable inference is that an identical limiting state of stress existed at each point on this surface. A similar region has been observed by Nadai[4] in the case of punching operations. The total volume of material in the heavily worked region increases from 0.03 in.^{3} for the $\frac{1}{2}$-in. diameter explosive to about 1.67 in.^{3} for the 2-in. diameter explosive.

Complete hardness surveys have been made of the specimens with a Rockwell hardness tester. The hardness of the material within the highly worked region was found to be very much greater than the hardness of the material just outside of this region. It changes abruptly as the boundary is crossed. Contours of equal hardness have been traced on the plates in Fig. 5. The contours approximate ellipses and bear a close resemblance to the isochromatics obtained in photoelastic studies.[5] An exact correspondence between the two cannot be established until more is known about the nature of the stress produced by the explosive charge.

The microstructures of the sections have also been studied. The parent material is shown in the upper left of Fig. 6. It is seen to be typical of an annealed medium carbon steel. The microstructure of the material below the high worked region, shown in the upper right photograph of Fig. 6, is characterized by the presence of a large number of Neumann bands or shock twins. These are known to occur on rapid loading[6] or at low temperatures under normal rates of loading.[7] The intensity of shock twinning was found to decrease with increasing distance from the explosive. Present evidence suggests that shock twinning is the principal mechanism responsible for the observed increase in hardness.

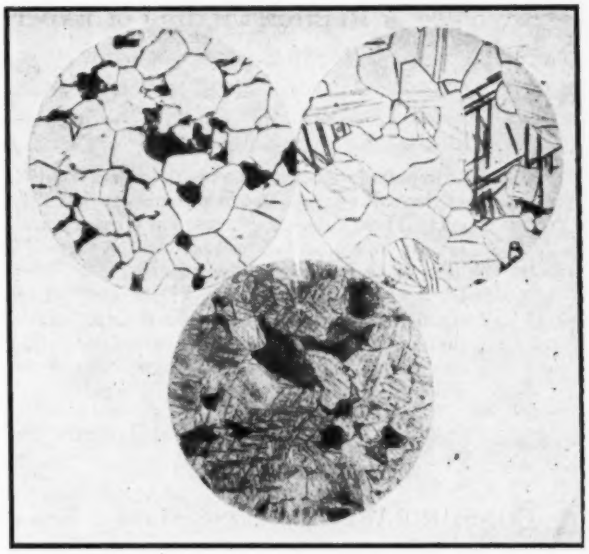


FIG. 6. Photomicrographs of sections. Upper left: parent material. Upper right: material below highly worked region. Lower center: highly worked material.

A typical photomicrograph of the material within the highly worked region is reproduced in the lower photograph of Fig. 6. The exact nature of the work-hardening mechanisms that are operative here is not well understood and further studies are now under way. It appears probable that extensive shock twinning occurs, but that superimposed on it is extensive slip and some actual breakup of the grains into crystallites.

CONCLUSIONS

It has been possible, by examining the several inelastic deformations produced in a block of steel, to establish qualitatively certain features of the stress distribution generated by the detonation of an explosive charge on the surface of the block. Generally, the stress pattern appears to correspond rather closely to that which would be produced by a static load except that shock twinning and fracturing both occur in the dynamic case because of the high rates of strain involved.

In conclusion the author wishes to acknowledge many helpful discussions with Mr. C. C. Woolsey and Dr. W. O. Wetmore.

[3] P. W. Bridgman, *Fracturing of Metals* (Am. Soc. Metals, 1948), p. 246.
[4] A. Nadai, *Plasticity* (McGraw-Hill Book Company, Inc., New York, 1931), p. 249.
[5] M. M. Frocht, *Photoelasticity* (John Wiley and Sons, Inc., New York, 1941), Vol. I, Fig. 10.16.
[6] Smith, Dee, and Young, Proc. Roy. Soc. (London) **A121**, 477 (1928).
[7] G. W. Geil and N. L. Carwile, J. Research Natl. Bur. Standards **45**, 129 (1950).

A Reinterpretation of Experiments on Intermetallic Diffusion*

A. S. Nowick†
Institute for the Study of Metals, University of Chicago, Chicago, Illinois
(Received April 9, 1951)

Reported values for the heat of activation, H, for volume chemical diffusion, as obtained from the slope of the best straight line in a semilogarithmic plot of diffusion data against T^{-1}, may often be greatly in error. These values will be too low whenever diffusion along internal surfaces plays a significant role at the lower temperatures of measurement. A method is presented which attempts to avoid such errors in H by using the calculated values of Zener for the intercept of the straight line. When the data is reconsidered in the light of this method, the widely accepted belief that H for chemical diffusion in dilute solid solutions is generally considerably less than H for self-diffusion of the solvent is brought to question. In fact, it is found that the chemical diffusion activation energy is generally within 15 percent of that for self-diffusion. It is further concluded that the process of diffusion under a chemical gradient may itself produce internal surfaces which act as short-circuiting diffusion paths.

INTRODUCTION

A CONSIDERABLE amount of effort has been expended in the acquisition of data[1] to describe the interdiffusion of two metals that form substitutional alloys. In accordance with custom, as well as theoretical expectation, the diffusion coefficient, D, is expressed according to an Arrhenius law,

$$D=D_0e^{-H/RT}, \tag{1}$$

whether or not the range and precision of the data are sufficiently good to define the parameters in such a law. It has become customary to tabulate D_0 and H for various diffusing systems, an excellent procedure when the data are precise and when a plot of log D *versus* T^{-1} gives a good straight line. Under other circumstances such a tabulation may be exceedingly misleading. In fact, there are many examples for which values of D at a given temperature, reported by different investigators, are in fair agreement but the values of D_0 and H reported differ considerably. For example, the self-diffusion of Cu, measured comparatively precisely with radioactive isotopes, has been reported[2] ranging from $D_0=0.1$ cm^2 sec^{-1}, $H=45$ kcal/mole to $D_0=47$ cm^2 sec^{-1}, $H=61.4$ kcal/mole. The reported values of D_0 and H should, therefore, not be taken too seriously; any correlations or conclusions involving these quantities may be regarded with suspicion. One widely accepted[3,4] conclusion is that the heat of activation, H, involved in the chemical diffusion of a dilute solute is considerably less than that for self-diffusion of the pure solvent. A second example is a correlation recently proposed by Dienes.[5] Both of these claims, based upon tabulated values of D_0 and H, will be brought to question in this paper.

It is generally assumed that the measured diffusion coefficients represent volume diffusion, i.e., the movement of atoms through the bulk of the crystal lattice rather than along grain boundaries or other internal surfaces. It is well known, however that atom movements along internal surfaces can and do occur.[6] Such effects are difficult to recognize from the diffusion data, except when these data cover a sufficiently wide range of temperature and are sufficiently precise to show a decided deviation from the straight-line semilog plot, with values of D abnormally high at low temperatures. Another criterion for the occurrence of diffusion along internal surfaces is an abnormally low value of H, but one must first know what values are "normal."

There are two questions to be answered if most of the data now available are to be of value: (1) How are we to know if the measurements really represent volume diffusion? (2) How are we to obtain a reasonable value for H when the range and precision of the data are not very good? A recent paper by Zener[7] suggests an answer to these questions, by deriving a theoretical equation from which reliable values of D_0 for volume diffusion can be predicted. The derived equation is in excellent agreement with the best data for self-diffusion; a similar equation has been used successfully in the case of interstitial diffusion.[8] The argument is as follows: It is readily shown that for either a vacancy mechanism or for the rotation of a ring of atoms, D_0 is given by

$$D_0=\gamma a^2\nu e^{\Delta S/R}, \tag{2}$$

* This research was supported in part by the ONR (Contract No. N6-ori-20-IV, NR 019 302).

† Present address: Hammond Metallurgical Laboratory, Yale University, New Haven, Connecticut.

[1] An excellent collection of diffusion data appears in C. J. Smithells, *Metals Reference Book* (Interscience Publishers, Inc., New York, 1949), pp. 390–412; and in W. Seith, *Diffusion in Metallen* (Verlag. Julius Springer, Berlin, 1939).

[2] Steigman, Shockley, and Nix, Phys. Rev. **56**, 13 (1939); B. V. Rollin, Phys. Rev. **55**, 231 (1939); M. S. Maier and H. R. Nelson, Trans. Am. Inst. Mining Met. Engrs. **147**, 39 (1942); Raynor, Thomassen, and Rouse, Trans. Am. Soc. Metals **30**, 313 (1942).

[3] A. D. Le Claire, Chapter VII of B. Chalmers, *Progress in Metal Physics I* (Interscience Publishers, Inc., New York, 1949).

[4] F. Seitz, *Physics of Metals* (McGraw-Hill Book Company, Inc., New York, 1943), chapter XII; F. Seitz, Acta Cryst. **3**, 355 (1950).

[5] G. J. Dienes, J. Appl. Phys. **21**, 1189 (1950).

[6] I. Langmuir, J. Franklin Inst. **217**, 543 (1934); J. A. M. Van Liempt, Rev. trav. chim. **51**, 114 (1932); R. F. Mehl, Trans. Am. Inst. Mining Met. Engrs. **122**, 11 (1936); R. S. Barnes, Nature **166**, 1032 (1950); R. M. Barrer, *Diffusion in and through Solids* (Cambridge University Press, New York, 1941), Chapter VII.

[7] C. Zener, J. Appl. Phys. **22**, 372 (1951).

[8] C. Wert and C. Zener, Phys. Rev. **76**, 1169 (1949); C. Wert Phys. Rev. **79**, 601 (1950); C. Wert, J. Appl. Phys. **21**, 1196 (1950).

where a is the lattice parameter; ν, the vibration frequency; R, the gas constant; and γ a numerical coefficient determined by the geometry of the elementary atomic jumps. For f.c.c. lattices, with which this paper is concerned, γ may be taken equal to unity. The quantity ΔS is the entropy of activation, related to H and the isothermal work, ΔG, required to carry the elementary diffusion process into the saddle position, by

$$\Delta G = H - T\Delta S. \tag{3}$$

From the thermodynamical equation,

$$\Delta S = -d\Delta G/dT, \tag{4}$$

only positive values ΔS are anticipated; if it is further assumed that a large part of the work ΔG goes into lattice strain, then

$$\Delta S = -(\lambda H/\mu_0)d\mu/dT, \tag{5}$$

where μ is the shear modulus (μ_0 at $T=0$), and λ is a numerical coefficient. This equation is in agreement with the data for self-diffusion when $\lambda = 0.57$; it should likewise apply to chemical diffusion in dilute solid solutions with the values of μ_0 and $d\mu/dT$ for the pure solvent.

If we substitute the Debye frequency for ν, all the factors required for the calculation of the volume-diffusion coefficient, D, as a function of temperature (Eqs. (5), (2), and (1)) are known, except for the heat of activation, H. A single observation of D at any finite temperature makes possible the calculation of H, and therefore of the complete diffusion curve, if we are sure that this observation really represents volume diffusion. If diffusion were entirely a volume process, any one of the experimental values would be satisfactory in making this calculation of H, and the calculated diffusion curve would be consistent with the remaining data within their experimental error. On the other hand, reported values[1] of D_0 for chemical diffusion are generally several orders of magnitude lower than the theoretical; usually they correspond to negative values of ΔS, in conflict with all experience which indicates that $d\Delta G/dT$ (Eq. (4)) should be less than zero. Thus we are led to anticipate that diffusion is significantly along internal surfaces, or "short-circuiting paths," as Zener calls them. Under these circumstances the volume effects would be expected to predominate at high temperatures. The most reliable value of the heat of activation for volume diffusion, therefore, should be that calculated from the experimental value of D at the highest temperature of measurement. This method will be applied in the next section to experimental data for chemical diffusion in dilute solid solutions, to obtain the most reliable values for the heat of activation for volume diffusion.

It may be argued that the assumptions made in deriving the theoretical equation for ΔS, and therefore for D_0, are not exact and, therefore, that it is not proper to force the intercept of the diffusion curve to go through such a value. The answer to this argument is that no attempt is being made to regard the theoretical D_0 as absolutely correct. Even if it were correct only in order of magnitude, the calculated value of H would be much more reliable than that obtained by drawing a straight line through the diffusion data, which can result in values of D_0 deviating from the theoretical by 10^8. The reason for this inability to determine reliable values H and D_0 from experimental data lies partly in the fact that most diffusion measurements are limited to between 2 and 3 logarithmic cycles in D, and partly in the fact that observations do not represent only volume diffusion. Further confirmation will be given to the viewpoints presented in this paper by a demonstration that the correlation claimed by Dienes[5] can be completely explained in terms of these considerations.

EXAMINATION OF THE DATA

An examination of tabulated values[1] of D_0 and H for various solutes in any single solvent metal shows that low values of H and small D_0 occur simultaneously, almost without exception. This observation is the basis of the correlation reported by Dienes,[5] in which the reported entropy of activation, $\Delta S'$, obtained from tabulated values of D_0, is shown to be a linear function of H'/T_m. Here H' is the tabulated heat of activation and T_m, the melting point of the solvent. The slope of the line is approximately unity, although the scatter of the data is quite large. Employed in this correlation are data for self-diffusion and for chemical diffusion of various solutes at low concentrations, primarily in Ag, Cu, and Pb. The use of primes in connection with the quantities $\Delta S'$ and H' expresses our doubt in their correctness, since in most cases D_0 is several orders of magnitude lower than the corresponding theoretical value. Let the true values, corresponding to the elementary volume diffusion process, be H and ΔS. The most reliable value of D is probably the one at the highest temperature of measurement, T_1. At this temperature, either the primed or unprimed set of parameters should give the same answer, or

$$\Delta S' - H'/T_1 = \Delta S - H/T_1. \tag{6}$$

Inasmuch as the highest temperature of measurement is always close to the melting point, T_m may be substituted for T_1 in Eq. (6). We have, therefore,

$$\Delta S' = (H'/T_m) + (\Delta S - H/T_m), \tag{7}$$

which is Dienes' correlation provided that $\Delta S - H/T_m$ is a constant for all materials. Actually, it is not constant; if ΔS is calculated for Cu, Ag, and Pb from Eq. (5), and the heat of activation for self-diffusion substituted for H, the bracketed quantity of Eq. (7) is lower by about 5.6 kcal/mole/°K for Pb than for Ag and Cu. It is readily verified that in the correlation curve presented by Dienes all the points for diffusion

TABLE I. Results of reanalysis of diffusion data.

System	D_0 (theoretical) $cm^2\ sec^{-1}$	H (calculated) kcal/mole	H (reported) kcal/mole	D_0 (reported) $cm^2\ sec^{-1}$
Ag in Ag	0.89	46	45.9	0.895
Cu in Ag	1.2	49	24.8	5.95×10^{-5}
Sb in Ag	0.73	44	21.7	5.31×10^{-5}
Sn in Ag	0.63	42.5	21.4	7.82×10^{-5}
Cd in Ag	0.51	41	24.8	7.3×10^{-5}
In in Ag	0.46	40	22.4	4.85×10^{-5}
Zn in Ag[a]	0.43	39	...	...
Cu in Cu	0.27	47	45 to 61.4	0.1 to 47
Pt in Cu	0.41	52.5	21.9	1.02×10^{-4}
Ni in Cu[a]	0.40	52	29.8	6.5×10^{-5}
Pd in Cu[a]	0.38	51.5	21.9	0.16×10^{-5}
Mn in Cu[a]	0.30	48	23.2	0.72×10^{-5}
Au in Cu[b]	0.23	...	44.9 ± 1.3	0.1 ± 0.06
Al in Cu	0.22	44.5	37.7	1.75×10^{-2}
Si in Cu	0.21	43.5	40	3.7×10^{-2}
Zn in Cu	0.21	43.5	...	...
Be in Cu	0.20	43	28	2.32×10^{-4}
Sn in Cu	0.17	40.5	45	1.13
Cd in Cu	0.15	39	8	1.97×10^{-9}
Pb in Pb	2.7	27	27.9	6.6
Ni in Pb	2.7	27	25.3	0.66
Sn in Pb	2.1	26	26.2	3.96
Tl in Pb	1.7	25	19.4	0.025
Bi in Pb	1.5	24.5	18.4	0.018
In in Pb	1.3	24	...	...
Mg in Pb	0.81	22	...	...
Cd in Pb	0.65	21	15.4	1.8×10^{-3}
Cu in Al	0.45	33	32.6	...[c]
Ag in Al	0.38	31.5	32.6	...
Si in Al	0.36	30.5	30.6	...
Mg in Al	0.32	30	32.2	...
Zn in Al	0.32	30	27.8	...

[a] In these cases the solute concentrations were relatively high, in the range 5 to 11 atomic percent.
[b] A. B. Martin and F. Asaro, Phys. Rev. **80**, 123(A) (1950).
[c] Some of the values of D_0 originally reported by Beerwald (Z. Elektrochem. **45**, 789 (1939)) are inconsistent with his data and the values of H he reports; they are therefore omitted here.

in Pb fall below those for Ag and Cu and by an amount close to the above calculated value. It appears, therefore, as if the correlation reported is simply a result of the fact that straight lines which do not correctly represent volume diffusion will simultaneously yield low D_0 and low H.

With this reassurance, we may now begin the program outlined in the "Introduction"—to calculate the most reliable value of H from the reported diffusion data by using the theoretical D_0 and the value of D at the highest temperature of measurement. In this manner we can obtain the most reliable straight line to describe *volume diffusion*. Large deviations from this straight line at the lower temperatures of measurement may then be regarded as evidence that short-circuiting paths play an important role in the diffusion experiments. The results are given in Table I; the data correspond to dilute solid solutions in Ag, Cu, Pb, and Al and to self-diffusion of the first three of these. Unless otherwise stated, the data are taken from references 1. The calculated values of H are expressed only to ±0.5 kcal because the data do not justify a closer calculation.

The most striking result is that H for chemical diffusion, when obtained by this method, is never far removed from the heat of activation for self-diffusion. In fact, in almost all cases of diffusion in Ag, Cu, and Pb the values of H are within 15 percent of that for self-diffusion. This result is in sharp contrast to the widely expressed belief[3] that relatively small values of H may be found in chemical diffusion, based upon the usual reported values listed in Table I. It should be emphasized that the calculated value of H is rather insensitive to the formula for D_0 and to the value of D used in its computation; errors in the measured D or in the theoretical value of D_0 up to a factor of 10, have only a small effect on the result. For example, values of D_0 obtained from the well-known Langmuir-Dushman equation[9] range from 0.15 to 0.7, so that if H were calculated using this equation and the high temperature data, the results would not be far different from those obtained using Eqs. (2) and (5). On the other hand, the reported values of H are strongly dependent upon the precision of the experiments. For illustration, the data of Rhines and Mehl[10] for diffusion in solid solutions of Cu are given in Fig. 1; the straight lines are those obtained from the calculated D_0 and H. The data for Cd in Cu, included in Fig. 1, deserve special mention; the tabulated value[1] of $H=8$ kcal/mole is sometimes taken seriously[5] in spite of the statement by Rhines and Mehl that the data may be regarded as correct only in order of magnitude. Because of the low temperatures at which these data were obtained, it is quite likely that the contribution made by short-circuiting paths is very significant.

Abnormally rapid diffusion has been observed[11] for Au and Ag into Pb; the data have been omitted from Table I, since the very small solubility in these cases suggests that diffusion may be almost entirely along internal surfaces. It has also been suggested[12] that diffusion of Au and Ag in Pb may be interstitial.

The values of H and D_0 for the diffusion of radioactive Au in Cu have been reported by Martin and Asaro and are listed in Table I. Although the actual data have not been given, the precision reported shows that, within the very small experimental error, D_0 is almost in perfect agreement with the theoretical value of 0.23 for Cu, while H is almost identical with the best data for self-diffusion of copper. Apparently, measurements of chemical diffusion, in which dilute radioactive

[9] S. Dushman and I. Langmuir, Phys. Rev. **20**, 113 (1922).
[10] F. N. Rhines and R. F. Mehl, Trans. Am. Inst. Mining Met. Engrs. **128**, 185 (1938).
[11] These data are summarized in references 1.
[12] W. Seith and A. Keil, Z. physik. Chem. **B22**, 350 (1933).

tracers are employed, can be made very precisely; further data of this kind will be very welcome.[13]

The straight lines obtained with the calculated values of H and D_0 often agree with the measurements within experimental error. In numerous systems, however, the measured values show large deviations from the calculated straight line at low temperatures; in these cases short-circuiting paths may contribute appreciably to the measured diffusion. Noteworthy among such cases are the data of Seith and Peretti[14] on the diffusion of Cd and In in Ag in which a break in the diffusion curve is obtained. The high temperature data are in fair agreement with the calculated D_0 and H of Table I, and are clearly representative of volume diffusion. Although it has been often indicated that the low temperature data must represent a short-circuiting process, the tabulated[1] D_0 and H are still generally obtained from an average straight line representative of all the data. Such tabulated data are highly misleading.

Whenever short-circuiting paths are the dominant factors in diffusion at low temperatures, they may still represent a significant part of the diffusion at the highest temperature of measurement. Under such circumstances, the heat of activation calculated by the method of this paper would be too low. Thus, an examination of Table I may lead one to conclude that perhaps if the data were precise enough and represented only volume diffusion, the heats of activation in self-diffusion and in chemical diffusion would be closer than 15 percent, perhaps even identical. Although at present one cannot decide from the data whether or not such a statement is correct, there exists a correlation that may have a bearing upon the subject. Köster and Rauscher[15] have reported the influence of small quantities of solutes on the elastic modulus of Cu, Ag, and Au. For alloys of Cu, the following elements *reduce* the elastic modulus: Sn, Be, Si, Mn, Zn, Al, Au (in order of decreasing magnitude of the effect), while Pd, Pt, and Ni *increase* the elastic modulus. Comparison with H-values (calculated) of Table I, shows that for solutes which decrease the modulus of Cu, H is lower than for self-diffusion (in almost the same order as the effect upon the modulus) and for those which increase the modulus, H is higher. The data for Mn (which seems to be an exception) are poor, and a lower H is quite possible. There are insufficient data on elastic moduli of solutions of Ag, Pb, or Al to attempt to extend this correlation to these cases. Two explanations for the agreement between elastic properties and heat of activation for diffusion are possible: (1) The presence of a finite amount (of the order of 2 to 8 percent) of solute is enough to change the elastic properties of the whole medium, with a corresponding change of the heat of activation (since the production of an elementary diffusion jump involves elastic distortion). (2) The factors responsible for change of modulus as solute is added (e.g., strain energy[16]) are also responsible for a corresponding change of the local environment of a solute atom so as to make a diffusion jump more or less difficult. If the first explanation is correct, the agreement between H for chemical diffusion and H for self-diffusion should improve as the solute concentration approaches zero. It should be pointed out that there is no theoretical reason, from the point of view of any of the proposed mechanisms of diffusion, for H to be exactly the same in chemical and self-diffusion.

The self-diffusion of Al has not been measured thus far. In view of the close agreement between values of the heat of activation for chemical and self-diffusion for solid solutions of Ag, Cu, and Pb, and in view of the close agreement between all calculated values of H for chemical diffusion in Al (see Table I), it appears that an excellent guess for the self-diffusion of Al would be

$$D=0.45\exp(-33000/RT)\ \text{cm}^2\ \text{sec}^{-1},$$

where $H=33\pm2$ kcal/mole.

DISCUSSION

On the basis of the present reanalysis of the data, we may conclude that tabulated values of H and D_0 for diffusion, as obtained from the best straight line through the experimental data, may be misleading. These values are very sensitive to experimental errors and are greatly influenced when low temperature diffusion data are, in a large part, a result of short-circuiting diffusion paths and not of volume diffusion. The method for the determination of H which is presented in this paper, attempts to avoid these difficulties and to obtain more reliable values of H than those reported in the literature. The results of the application

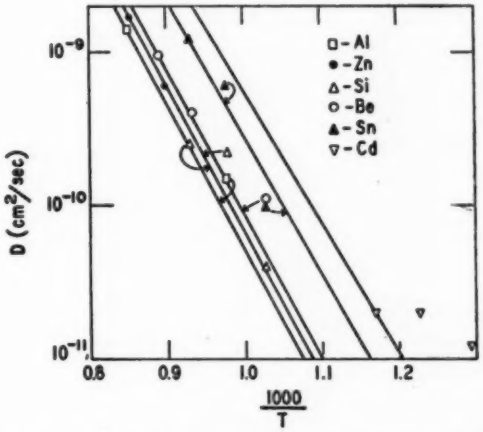


FIG. 1. The data of Rhines and Mehl for the chemical diffusion of low concentrations of various solutes in copper. The straight lines are obtained from the calculated values of D_0 and H.

[13] Data for the diffusion of radioactive Co into Ni have been reported by R. C. Ruder and C. E. Birchenall, J. Metals (February, 1951), as $D_0=1.46$ cm² sec⁻¹ and $H=68300$ cal/mole. The theoretical D_0 for Ni is about 2.6, so that the agreement here is excellent.

[14] W. Seith and E. Peretti, Z. Elektrochem. **42**, 570 (1936).

[15] W. Köster and W. Rauscher, Z. Metallk. **39**, 111 (1948).

[16] C. Zener, Acta Cryst. **2**, 163 (1949).

of this method suggests that, at least in f.c.c. lattices, the heat of activation for chemical diffusion in dilute solid solutions is much closer to the values of H for self-diffusion than had been hitherto suspected.

It is interesting to note that values of D_0 for self-diffusion in cubic materials, as measured by radioactive tracers, are generally quite close to theoretical values, in contrast to the data for chemical diffusion. This fact suggests that the role of short-circuiting paths may not be as great in self-diffusion as in chemical diffusion. The samples used in self-diffusion experiments are usually polycrystalline and no less imperfect at the start of the experiments than those in chemical diffusion experiments. It seems, therefore, that, in order to explain the apparently greater importance of nonvolume mechanism in chemical diffusion experiments, we are led to one of the following conclusions: Either (1) solute atoms have a greater preference for short-circuiting paths (such as grain boundaries present in the sample) than the solvent atoms, or (2) diffusion under a chemical gradient produces *new short-circuiting paths* because of the strains involved in the gradient. If (1) is true, special care must be taken in studying chemical diffusion to avoid internal surfaces, i.e., samples should be well-annealed single crystals. On the other hand, if (2) is the true explanation only chemical diffusion experiments involving negligibly small gradients and similar (in chemical properties and in size factor) elements will yield high precision data relating to the volume diffusion process. Thus, Johnson[17] has measured the mutual chemical diffusion of Ag and Au, which are very similar in size and properties, near the 50–50 composition with only a 4-percent concentration gradient. His precision is very good, and the value of D_0 obtained very close to the theoretical, in contrast to earlier data (which he reviews) in which larger gradients were used. In the latter, the usual very small D_0 and small H were observed.

It has been shown that in some chemical diffusion experiments recrystallization occurs in the diffusion zone.[10,18] Thus, new internal surfaces have been formed, as described by (2). In other experiments in which the strains produced by the chemical gradient are less severe, it is still conceivable that new short-circuiting paths, such as dislocations, are produced, even though no recrystallization is observed. Whichever of the possibilities, (1) or (2) already described, may apply to most experiments, the fact remains that special care is needed to obtain purely volume diffusion. If diffusion is along short-circuiting paths, there is no *a priori* reason to believe that the two components will diffuse at the same rates. Thus one may observe movements of inert markers, the so-called Kirkendall effect, resulting from unequal diffusion of the two components. In experiments[19] in which such marker movements have been observed, large concentration gradients are used, so that the possibility that the marker movements are partially or completely the results of nonvolume diffusion effects certainly must be taken into account. This possibility is recognized by da Silva and Mehl.[19] On the other hand, in view of the writer's recent demonstration[20] that an excess of vacancies may be quenched into an alloy, it is quite probable that diffusion in f.c.c. lattices is by means of vacancies. It follows that the diffusion of the solute and solvent atoms are not necessarily at the same rate and, hence, that at least part of the marker movements may be a result of purely volume diffusion.

In conclusion, the author wishes to thank Professor C. Zener for helpful conversations.

[17] W. A. Johnson, Trans. Am. Inst. Mining Met. Engrs. **147**, 331 (1942).

[18] See, for example, reference 6, R. F. Mehl.

[19] A. D. Smigelskas and E. O. Kirkendall, Trans. Am. Inst. Mining Met. Engrs. **171**, 130 (1947); L. C. C. da Silva and R. F. Mehl, J. Metals, Feb., 1951.

[20] A. S. Nowick, Phys. Rev. **82**, 551 (1951).

JOURNAL OF APPLIED PHYSICS VOLUME 22, NUMBER 9 SEPTEMBER, 1951

The Generation and Measurement of Low Frequency Random Noise

R. R. BENNETT AND A. S. FULTON
Hughes Aircraft Company, Culver City, California
(Received April 9, 1951)

Judgment of the performance of automatic control systems often involves a consideration of the behavior of such systems in the presence of random noise. The noise may be of an undesirable nature, such as receiver noise, or it may represent an ensemble of statistical inputs which the system is to follow. Analytical techniques exist for determining the response of linear systems to noise inputs. However, no general methods are available for analyzing nonlinear systems subject to noise inputs. An electric analog computer, together with a suitable source of noise, may be used to study nonlinear systems. This paper describes an electronic noise generator designed for analog computer or simulator use. Methods are discussed for measuring the important characteristics of low frequency noise, such as the mean value, spectral density, amplitude distribution, and autocorrelation. Particular attention is devoted to the length of time necessary to establish satisfactory estimates of the properties of low frequency noise.

I. GENERATION OF NOISE

WE shall concern ourselves with a particular type of noise, namely, a continuous stationary time series. Such a noise will have two fundamental characteristics: an amplitude distribution, which specifies the relative probability that a given noise amplitude will occur; and a power spectrum which describes the division of the noise power into frequency components.[1] The noise generator to be described has an output characterized by a normal or gaussian amplitude distribution and a power spectrum that is essentially uniform from zero to 25 cycles per second. The band width of the noise could be extended at the expense of a correspondingly lower noise power density at the low frequencies.

A. Description of Noise Generator

Figure 1 shows a block diagram of the noise generator. A Type 2D21 gas thyrathron is used as the primary noise source. The cathode and grids of the thyratron are connected together, and a plate load resistance of 0.1 megohm is connected from the plate to the plus 300-volt supply voltage.

A narrow band of the thyrathron output, centered at 4 kc, is passed by a selective amplifier having a single parallel *LC* tuned circuit. The *Q* of the toroidal inductor used is approximately 100, and the resultant band width of the noise (measured between half-power points) is approximately 50 cycles per second. The output of the tuned amplifier is then transformer-coupled to a ring demodulator, as shown. The demodulator is also supplied with a 4-kc square wave. The amplitude of the square wave is large compared with the noise, and hence the demodulator serves to multiply the noise voltage by the square wave. This multiplication results in an output of sum and difference frequencies having a low frequency component of frequency Δf corresponding to a component of frequency 4 kc plus or minus Δf at the tuned amplifier output, together with signals centered at 4 kc and harmonics thereof. The condenser, *C*, across the output serves to attenuate the high frequencies, leaving only the low frequency noise.

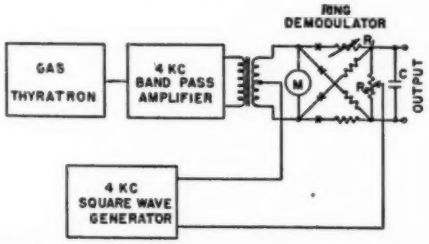


FIG. 1. Block diagram of random noise generator.

The ring demodulator uses a Type 1N42 crystal diode (4 matched diodes in one envelope). Each of three diodes has a 500-ohm resistor connected in series, and the fourth has a rheostat, R_1, used for balancing the demodulator. A second balance is effected by potentiometer R_2.

A meter (*M*, Fig. 1) is used to indicate the noise generator output by measuring the output of the tuned amplifier. This meter is an averaging type having a time constant of several seconds.

It is essential that the frequency of the square wave be accurately centered at the center of the tuned amplifier pass band. For this reason a stabilized core inductor and low (or negative) temperature coefficient capacitors are used in the tuned circuit of the amplifier. In addition, the square wave frequency is controlled by a sealed driven tuning fork.

B. Nature of Noise Generator Output

A recorded sample of the output of the noise generator is shown in Fig. 2. The rms output is approximately 5 volts. The demodulator will maintain a balance over

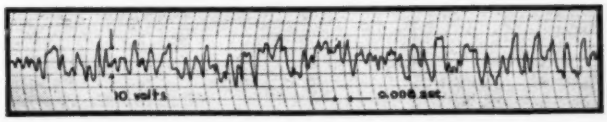


FIG. 2. Recorded sample of output of random noise generator.

[1] S. Goldman, *Frequency Analysis, Modulation and Noise* (McGraw-Hill Book Company, Inc., New York, 1948).

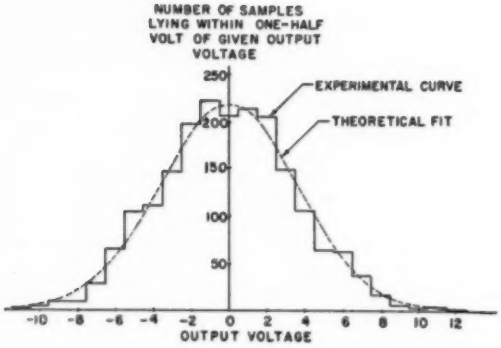


FIG. 3. Histogram of output of random noise generator.

extended periods of time and over normal variations in room temperature, such that the mean value of the output remains less than 0.05 volt.

From an extended record of the type shown in Fig. 2 a histogram has been plotted as the solid curve of Fig. 3, showing the probability density of occurrence of a given output voltage versus the value of the voltage. This plot represents a total of 1976 sample points of the noise record. The continuous dotted bell-shaped curve is a "best fit" gaussian distribution curve calculated according to the theory of statistical estimation.[2] A χ^2 (chi-squared) test indicates that the hypothesis that the experimental distribution is a normal distribution is substantiated. (For the indicated theoretical gaussian distribution, the probability of obtaining a χ^2 greater than that for this sample is 0.083, where there are 20 degrees of freedom in the sample. Ordinarily, statisticians consider a theoretical distribution to be substantiated if, as a result of a χ^2 test, the foregoing probability is greater than 0.05.)

The output spectrum has been checked by measurement methods to be described later, and it has been verified that the spectrum is just that determined by the band pass characteristic of the 4-kc tuned amplifier (that is, a uniform spectrum at low frequencies with a half-power point at approximately 25 cps).

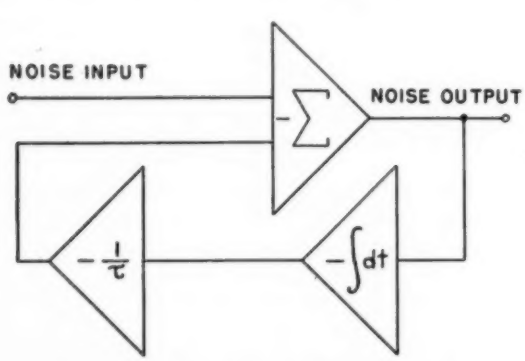


FIG. 4. Circuit for removing mean of noise.

[2] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, New Jersey, 1946).

C. Removing the Mean Value

It is often desirable to have a source of noise of zero mean output. The mean output of the generator described herein, while sufficiently low for most simulator purposes, can be made negligibly small by utilizing an assemblage of analog computer amplifiers to form a high pass filter. Figure 4 shows such a filter where the operation performed by each amplifier is indicated on its symbol. A mean value or dc component in the input will be canceled by an equal and opposite voltage fed back from the integrator. The mean value of the output is then equal to the equivalent drift voltage at the integrator input, which can be made less than 10 microvolts. The low frequency cutoff of the filter, determined by τ as shown, can be made arbitrarily low consistent with the value of the input noise and the dynamic range of the integrator.

D. Noise with Other Characteristics

Beginning with a noise generator whose output spectrum is uniform over a sufficient range, it is a simple matter to alter the spectrum shape in a known manner by passing the noise through filters composed of passive elements or feedback amplifiers. Indeed, one may even generate special types of nonstationary time series by utilizing time-varying filters. Noise having certain distributions other than gaussian may be generated from gaussian noise by combinations of linear and nonlinear filtering. For example, narrow band filtering and subsquent envelope detection of gaussian noise results in an output having a Rayleigh distribution.

II. MEASUREMENT OF NOISE

Perhaps the most demanding and least appreciated aspect of low frequency noise measurement is the length of measurement time necessary to establish accurate estimates of the statistical properties of the noise. For example, to determine the spectrum shape at an arbitrarily low frequency requires an observation time which may become prohibitively long. This is not a limitation imposed by technique but rather a fundamental limitation. For this reason a large part of the discussion to follow is concerned with the length of time required for an observation.

A. Periodic Components

If it cannot be accepted *a priori* that a given source of continuous noise is indeed random and contains no periodic components, a measurement can be made to determine the possible presence of such periodicities. (Strictly speaking, one can determine only that the periodic components, if present, are "unlikely," in a statistical sense, to be greater than an arbitrarily small amount.) Perhaps the most direct test for hidden periodicities is by means of the autocorrelation func-

tion.[3] While the autocorrelation function can be computed automatically by an analog computer, it is felt to be just as simple to record a sample of the noise and determine it by a combination of sampling and numerical computation.

B. Determination of Mean Value

The mean value of a noise voltage may be found by integrating the voltage for a sufficiently long time. The output of the integrator will consist of a component which increases linearly with time, due to the mean of the input, plus a fluctuation due to the ac components of the input noise. After sufficient elapsed time the ac component fluctuation will be small compared with the contribution of the mean, and consequently an increasingly accurate estimate of the mean is obtained. It is demonstrated in the Appendix that if the noise is normally distributed, consists of a uniform spectrum of spectral density N_0 volts^2/cps at all frequencies other than zero, and has a true mean value $\bar{e}$ volts, it is necessary to continue the integration process for a length of time, T, given by

$$T=(2\times 10^4 N_0)/P^2\bar{e}^2 \text{ sec} \tag{1}$$

in order to insure with 95 percent certainty (that is, there is one chance in twenty, on the average, of being wrong) that the measured mean lies within P percent of the true mean.

If the spectral density of the noise is given by

$$w(f)=\frac{N_0}{1+(f^2/B^2)} \tag{2}$$

(that is, a band width of B cps) Eq. (1) is a valid approximation, as is shown in the Appendix, provided $BT\gg 1$.

Consider now the problem of determining within an accuracy of 10 percent the mean value of the output of the noise generator described previously. Let us assume the true mean value is 0.05 volt and that the spectral density, N_0, is 0.6 volt^2/cps. Then Eq. (1) prescribes that

$$T=\frac{2\times 10^4\times 0.6}{(10)^2\times(0.05)^2}=4.8\times 10^4 \text{ sec}$$

or over 13 hours of integration time is required. Such a calculation lends an appreciation of the lengthy periods involved in making such measurements.

C. Determination of Mean Square Value

An analog computer may be used to determine the mean square value, $\langle e^2\rangle$, of the noise by squaring the noise voltage and integrating this over a period of time. However, if the noise frequencies involved are high there is usually considerable difficulty involved in accurately computing the square. If it is known that the noise is normally distributed and has zero mean value, the mean square value may be inferred from the average of the positive (or negative) values of the noise voltage using the relationship

$$\langle e^2\rangle=2\pi A_+^2, \tag{3}$$

where A_+ is the average of the positive values of the noise. A_+ can be determined by half-wave rectifying the noise and then integrating. This method is essentially unlimited in terms of frequency response and can be made to yield considerable accuracy.

As was the case in determining the mean value, there is necessarily a minimum time of observation associated with determining the mean square (or the mean square deviation, defined as $\langle e^2\rangle-\bar{e}^2$, if the noise has a mean value). It is shown in the Appendix that the length of time required to determine, with 95 percent certainty, the mean square deviation to within a percentage accuracy P, for normally distributed noise having a spectral density given by Eq. (2), is approximately

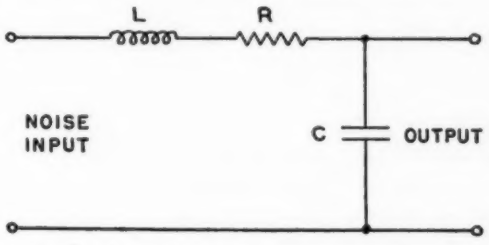


FIG. 5. Circuit for simple spectrum analyzer.

$$T=(4\times 10^4)/\pi BP^2 \text{ sec.} \tag{4}$$

As would be expected, the time required is inversely proportional to the band width.

D. Spectrum Analysis

There are various methods for measuring the shape of the spectrum of a particular noise. One method often used involves a measurement or computation of the autocorrelation function and subsequent determination of the power spectrum from it. Since in simulator work the noise source is usually continuously available, it is most often simpler, particularly if the spectrum is known to be reasonably uniform in the regions of interest, to pass the noise through selective filters and then measure the transmitted power. Repeating this process at a number of different frequencies will give a measure of the spectrum shape. For example, feedback amplifiers may be connected to form the analog of the simple selective filter shown in Fig. 5. If the Q of this circuit* is made large, the output will indicate the spectral density of the input in the vicinity of the

[3] *Theory of Servomechanisms*, MIT Radiation Laboratory Series, Volume 25.

* See Appendix for definition of Q.

resonant frequency. For large values of Q the mean square deviation of the output is given approximately by

$$\langle e^2\rangle - \bar{e}^2 \approx \pi f_0 N_0 Q/2, \tag{5}$$

where f_0 is the resonant frequency in cycles per second.

N_0 is the input spectral density in the vicinity of f_0 in volts²/cps, and hence by measuring $\langle e^2\rangle - \bar{e}^2$ by the method cited previously, a measure of N_0 is obtained. For an accuracy of measurement (see Appendix) of the same certainty as that specified for the previous types of measurement, the time involved must be at least

$$T = (4\times 10^4 Q)/\pi f_0 P^2, \tag{6}$$

where $Q \gg 1$.

Filters of a more complicated nature may be constructed to have sharper selectivity properties. Let us consider as a limiting case a filter which has unity transmission for $f_0 < f < f_0 + \Delta f$ and zero transmission for all other frequencies. (This is not physically realizable but may be approached as a limit.) The output mean square value for this case is simply $N_0 \Delta f$, where N_0 is the input spectral density (assumed constant) in the vicinity of f_0. The time of measurement required is shown in the Appendix to be, approximately,

$$T = (4\times 10^4)/P^2 \Delta f \text{ sec} \tag{7}$$

provided $f_0 \gg \Delta f$.

For the circuit shown in Fig. 5 the band width may be considered to be the frequency difference between half-power points, which is approximately

$$\Delta f = f_0/Q$$

for large Q. Equation (6) then becomes, upon substitution,

$$T = (4\times 10^4)/\pi P^2 \Delta f,$$

which is of the same order of magnitude as Eq. (7).

It is to be noted that the time of observation required for selective filters of this type is inversely proportional to the band width. Hence, to obtain finer resolution of the spectra requires longer and longer observation times. This is analogous to the fundamental limitation imposed upon spectral measurements at frequencies near zero.

The foregoing expressions for observation time have been specified for normally distributed noise having particular spectra. However, they may be applied as first order approximations to cases involving other distributions and spectral shapes.

APPENDIX

The significant formulas presented in the text are developed here.

The integration of a stationary variable for a period of time T may be considered as a multiplication of that variable by a weighting function[3] which is unity for $0 < t < T$ and zero elsewhere. If the variable being integrated has a power spectral density $w(f)$, $f \neq 0$, the mean square deviation of the output at the end of time T is given by

$$\int_0^\infty w(f) S(f) df,$$

where

$$S(f) = \sin^2 \pi f T/\pi^2 f^2 \tag{8}$$

may be shown to be the spectral density of the weighting function.

The mean square deviation, σ^2, of the integral, after time T, of a noise voltage having a uniform spectral density N_0 is given by the above to be

$$\sigma^2 = \int_0^\infty \frac{N_0 \sin^2 \pi f T}{\pi^2 f^2} df = \frac{N_0 T}{2}. \tag{9}$$

The output values obtained from such integrations (that is, from the ensemble[3]) will be normally distributed if the original noise is so distributed, and hence 95 percent of them will, on the average, lie within $\pm 2\sigma$ of the true mean value (obtained after infinite time). Thus, for 95 percent certainty of a percentage accuracy P

$$0.01 P \bar{e} T = 2\sigma = 2(N_0 T/2)^{\frac{1}{2}}$$

or

$$T = (2\times 10^4 N_0)/P^2 \bar{e}^2, \tag{1}$$

which is the result given in the text.

For the spectral density given by Eq. (2) we obtain

$$\sigma^2 = \int_0^\infty \frac{N_0}{1+(f^2/B^2)} \cdot \frac{\sin^2 \pi f T}{\pi^2 f^2} df = \frac{N_0 T}{2}\left[1 - \frac{1}{2\pi BT}(1 - \epsilon^{-2\pi BT})\right],$$

which, for $BT \gg 1$, is approximately the same as Eq. (9).

To determine the mean square value of a noise voltage, we may proceed by determining first the spectral density of the square of the voltage. Let

$$I \equiv e^2$$

be the instantaneous value of the square. Then, if e has a spectral density $w(f)$, the spectral density of I is given by[4]

$$W(f) = \int_{-\infty}^{\infty} w(x) w(f-x) dx, \tag{10}$$

where

$$w(-x) = w(x)$$

and

$$f \neq 0.$$

The mean square deviation of $R \equiv \int_0^T I dt$ is given by

$$\sigma^2 = \int_0^\infty W(f) \frac{\sin^2 \pi f T}{\pi^2 f^2} df. \tag{11}$$

If e has a spectral density given by Eq. (2) the above yields

$$W(f) = \frac{\pi B N_0^2}{2[1+(f^2/4B^2)]}$$

and

$$\sigma^2 = \frac{T \pi N_0^2 B}{8}\left[2 - \frac{1}{2\pi BT}(1 - \epsilon^{-4\pi BT})\right]. \tag{12}$$

For $BT \gg 1$, which is the case of interest,

$$\sigma^2 \approx \pi N_0^2 BT/4.$$

The values of R will not be normally distributed. However, as σ becomes smaller and smaller with respect to $\bar{R}$ (the ensemble average), the distribution approaches a normal one, and it is sufficiently accurate to assert that 95 percent of the values of R lie within $\pm 2\sigma$ of $\bar{R}$. Now $\bar{R}$ is given by

$$\bar{R} = T \int_0^\infty w(f) df, \tag{13}$$

which, in this case, yields

$$\bar{R} = \pi N_0 BT/2.$$

Setting

$$2\sigma = 0.01 P \bar{R},$$

we obtain

$$T = (4\times 10^4)/\pi B P^2. \tag{4}$$

For the "spectrum analyzer" circuit of Fig. 5 the output spectral density corresponding to an input spectral density N_0 (as-

[4] S. O. Rice, Bell System Tech. J. (January, 1945).

sumed constant over at least a narrow range of frequency) is

$$w(f)=\frac{N_0Q^2}{Q^2[1-(f^2/f_0^2)]^2+(f^2/f_0^2)}, \tag{14}$$

where $Q\equiv 1/R(C/L)^{\frac{1}{2}}$. Now

$$\left(1-\frac{f^2}{f_0^2}\right)^2=\left(1-\frac{f}{f_0}\right)^2\left(1+\frac{f}{f_0}\right)^2.$$

For $Q\gg 1$ the value of $w(f)$ is much larger near $f=f_0$ than elsewhere, and for the subsequent integrations we may make the approximation

$$Q^2\left(1-\frac{f}{f_0}\right)^2\left(1+\frac{f}{f_0}\right)^2+\frac{f^2}{f_0^2}\approx 4Q^2\left(1-\frac{f}{f_0}\right)^2+1,$$

whereby Eq. (14) becomes, approximately,

$$w(f)=\frac{N_0Q^2f_0^2}{4Q^2(f_0-f)^2+f_0^2}. \tag{15}$$

Using Eqs. (13) and (15) we obtain for this case

$$\bar{R}=N_0f_0\pi QT/2. \tag{16}$$

To perform the integration of Eq. (10) we use

$$w^*(x)=\frac{N_0Q^2f_0^2}{4Q^2(f_0+x)^2+f_0^2},$$

whenever $x<0$.

The values of the integrand in Eq. (10) corresponding to the range of integration are then

$$\begin{array}{ll} -\infty<x<0 & w^*(x)w(f-x) \\ 0<x<f & w(x)w(f-x) \\ f<x<\infty & w(x)w^*(f-x) \end{array}$$

and we obtain

$$W(f)=\frac{N_0^2f_0^4}{16}\left\{\int_{-\infty}^{0}\frac{dx}{[(f_0+x)^2+(f_0^2/4Q^2)][(f_0-f+x)^2+(f_0^2/4Q^2)]}\right.$$
$$+\int_0^f\frac{dx}{[(f_0-x)^2+(f_0^2/4Q^2)][(f_0-f+x)^2+(f_0^2/4Q^2)]}$$
$$\left.+\int_f^{\infty}\frac{dx}{[(f_0-x)^2+(f_0^2/4Q^2)][(f_0+f-x)^2+(f_0^2/4Q^2)]}\right\}. \tag{17}$$

We are interested in only the low frequency values of $W(f)$, which clearly receive contributions from only the first and third integrals of Eq. (17). Designating by $W_L(f)$ the low frequency portion of $W(f)$, we have, after making a change of variable,

$$W_L(f)=\frac{N_0^2f_0^4}{8}$$
$$\times\int_{-\infty}^{0}\frac{du}{[(f_0+u)^2+(f_0^2/4Q^2)][(f_0-f+u)^2+(f_0^2/4Q^2)]}. \tag{18}$$

Again making use of the assumption $Q\gg 1$ we may extend the limits of integration of Eq. (18) without appreciably affecting the result, giving

$$W_L(f)\approx\frac{N_0^2f_0^4}{8}$$
$$\times\int_{-\infty}^{\infty}\frac{du}{[(f_0+u)^2+(f_0^2/4Q^2)][(f_0-f+u)^2+(f_0^2/4Q^2)]}$$
$$=\frac{N_0^2f_0\pi Q^2}{2[1+(Q^2f^2/f_0^2)]}. \tag{19}$$

Substituting the above value of $W_L(f)$ into Eq. (11) there results

$$\sigma^2=\frac{\pi N_0^2f_0Q^2T}{8}\left[2-\frac{Q}{\pi f_0T}\left(1-\exp\left(-\frac{2\pi f_0T}{Q}\right)\right)\right] \tag{20}$$

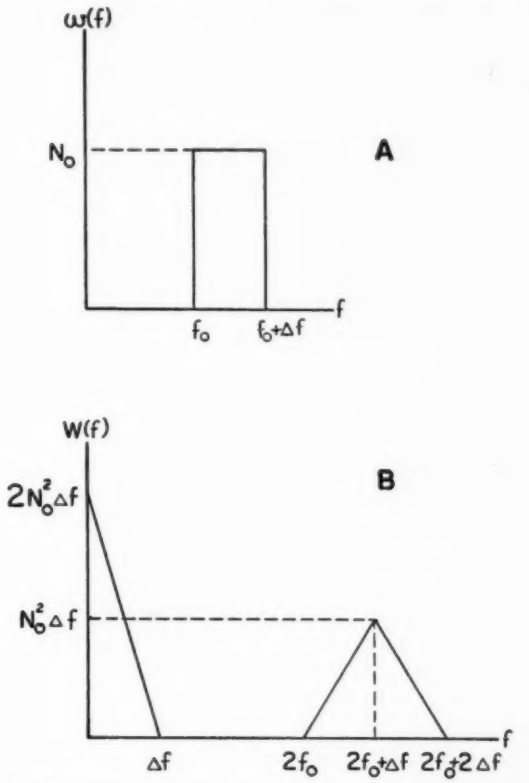


FIG. 6. (A) Assumed power spectrum of narrow band noise voltage. (B) Resultant power spectrum of square of noise voltage.

which, for $f_0T\gg Q$, becomes approximately

$$\sigma^2\approx\frac{N_0^2f_0Q^2\pi T}{4}. \tag{21}$$

For

$$2\sigma=0.01P\bar{R}$$

we get

$$T=(4\times 10^4Q)/\pi f_0P^2, \tag{6}$$

which is, incidentally, consistent with the foregoing assumption that $f_0T\gg Q$ (provided, of course, that P has a reasonable value).

For the sharp spectrum of Fig. 6A we obtain for $W(f)$, using Eq. (10), the form shown in Fig. 6B, for which, by Eq. (11)

$$\sigma^2=\int_0^{\Delta f}\frac{2N_0(\Delta f-f)\sin^2\pi fTdf}{\pi^2f^2}$$
$$+\int_{2f_0}^{2f_0+\Delta f}\frac{N_0^2(f-2f_0)\sin^2\pi fT}{\pi^2f^2}df$$
$$+\int_{2f_0+\Delta f}^{2f_0+2\Delta f}\frac{N_0^2(2f_0+2\Delta f-f)\sin^2\pi fT}{\pi^2f^2}df. \tag{22}$$

For $T\Delta f\gg 1$ and $f_0\gg\Delta f$ only the first integral (that is, the low frequency portion) in Eq. (22) contributes significantly. Hence we obtain, within the above approximation,

$$\sigma^2=N_0^2T\Delta f.$$

For this example

$$\bar{R}=N_0\Delta fT.$$

Setting

$$2\sigma=0.01P\bar{R},$$

we obtain

$$T=(4\times 10^4)/P^2\Delta f. \tag{7}$$

Carbon, Oxygen, and Sulfur Content of Chilean Coppers as Related to Cuprous-Oxide Rectifiers

C. C. Hein and W. M. Hickam
Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania

(Received April 12, 1951)

Several Chilean coppers used in cuprous-oxide rectifier work have been analyzed for carbon, oxygen, and sulfur. The carbon determinations were in agreement with published solubility data. Large differences were found in the oxygen and sulfur content which could be correlated with the refining data of the copper and the reverse leakage currents of the rectifiers. The change in sulfur content caused by various heat treatment also was investigated.

I. INTRODUCTION

In the manufacture of cuprous-oxide rectifiers the material used is almost exclusively Chilean copper. This copper is of much higher purity as obtained from the refinery than coppers from other parts of the world. The purity is of the order of 99.95 percent. The largest single impurity is oxygen, which averages approximately 0.03 percent. It is generally agreed that the reason only Chilean copper can be used for cuprous oxide rectifiers is because of its small impurity content.[1] However, the characteristics of cuprous-oxide rectifiers reveal large differences in Chilean coppers. It is the purpose of this paper to present the results obtained from the carbon, oxygen, and sulfur analyses of a large number of Chilean coppers, and to show how these results correlate with the rectifier characteristic.

The coppers included in this report were all Chilean coppers. They represent typical coppers received over the past ten years and were selected primarily on the basis of rectifier history. The refinery lot numbers given refer to the shipments of copper as received from Chile. The sheet or strip identification means the information was obtained from a single sheet or strip of the particular refinery lot. "Pan" identification refers to copper of a single refinery lot number with no selection as to the part of the lot. Refinery lot M-135 is divided into three subheads S, M, and E. This refers to the start, middle, and end of a single pour of copper at the refinery. The furnace is tapped from the bottom, so the start is the bottom of the melt and the end is the top of the melt.

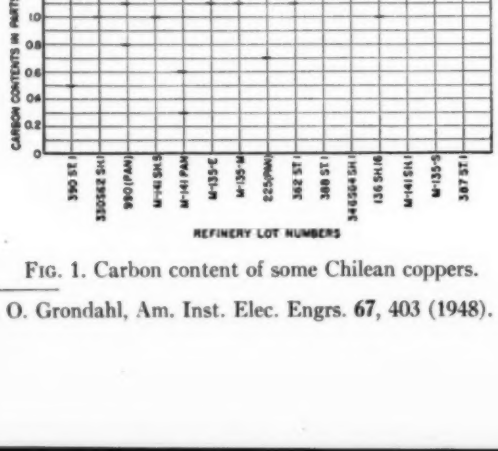


FIG. 1. Carbon content of some Chilean coppers.

The rectifier data presented in this report were obtained on standard 1½-in. o.d. washers. The results represent the average of sets of data obtained on rectifiers processed in identical fashion. The method of processing was as follows.

The washers were oxidized in tank oxygen at 1025°C for seven minutes. They were then cooled in nitrogen to 523°C and annealed for thirteen minutes, followed by quenching in distilled water.

Units consisting of six cells were made up and tested for one week at a total of 30 volts per unit in the reverse direction at 35°C. The reverse leakage currents reported are those obtained after one week of test. When data were available for more than one unit, the results were averaged. The processing method described was followed closely in obtaining the reverse leakage currents contained in this paper except where specified otherwise. These currents are a measure of the quality of the rectifier. Coppers with values less than 25–35 ma are, in general, considered satisfactory. The forward drops showed small variations but were not found to be as indicative of rectifier performance under the described processing as were the reverse leakages.

In order to emphasize the difficulty generally encountered in obtaining rectifier data which limits the information to that of a qualitative nature, a comparison shall be given for M-135 copper when processed in O_2 and N_2 as described earlier and when processed throughout in air.

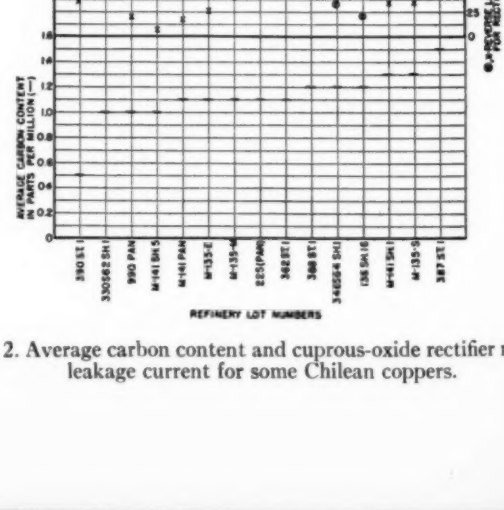


FIG. 2. Average carbon content and cuprous-oxide rectifier reverse leakage current for some Chilean coppers.

[1] L. O. Grondahl, Am. Inst. Elec. Engrs. **67**, 403 (1948).

Refinery lot number	Reverse leakage current (ma) O_2 and N_2 processing	Reverse leakage current (ma) Air processing
M-135-S	35.0	30.5
M-135-M	39.4	39.8
M-135-E	26.8	62

The effects of different processing and other variations encountered in rectifier production make it difficult to classify a particular copper unless a set procedure is developed and followed with great precision; even then the results must be interpreted on a statistical basis.

II. EXPERIMENTAL RESULTS

Carbon

The methods of analyses used for carbon, oxygen, and sulfur determinations have been described in an earlier paper.[2] Carbon analyses obtained from a number of Chilean coppers are shown in Fig. 1. Each point represents a single analysis. The refinery lot numbers have been arranged such that the average carbon content increases in going from left to right. In the small concentration studied it is difficult to say what fraction of the spread in the points is due to non-uniform carbon content and what fraction is due to experimental errors. The average of these results are shown in Fig. 2 along with the reverse leakage current in milliamperes. The average carbon content approximates one part per million by weight and is in agreement with the published solubility data of Beaver and Floe[3] for a refining temperature of 1100°C. No correlation is found between carbon content and reverse leakage current in this copper.

Oxygen

Oxygen, the largest single impurity in Chilean copper, is subject to control at the refinery by "poling." Because of its high concentration and the ease with which it is removed, it is an easy impurity to determine. In Fig. 3 are given the oxygen determinations for a number of Chilean coppers. The refinery lot numbers have been arranged so as to yield increasing average oxygen content in going from left to right. A range of some 210 to 380 parts per million was found in these coppers. M-141 is an interesting example of the spread that can occur in a single refinery lot. Sheet 1 contains approximately 30 percent more oxygen than sheets 3, 4, 5, and 6 of this refinery lot number. Refinery lot M-135 shows a similar variation. The spread in the analyses of 990 (pan) and 225 (pan) undoubtedly represent variations that occur in a single refinery lot.

A significant correlation is also found between refinery date and oxygen content. The coppers M-141, 136, 990, and M-135 are "old" coppers and date from 1940 to 1946. The refinery lots 346S64, 225, 330S62, 362, 387, 388, 389, and 390 are "new" coppers and date from 1946 to the present. In general, the "old" coppers contain less oxygen than the "new" coppers. In Fig. 4 are presented the average oxygen content and reverse leakage current for the various coppers. There is a strong correlation of oxygen content and reverse leakage current. The low oxygen content coppers yield the best quality rectifiers in general. It follows that the quality of the copper as far as rectifier work is concerned has deteriorated. A few exceptions are found from these general trends which indicate that oxygen may not be the complete source of trouble.

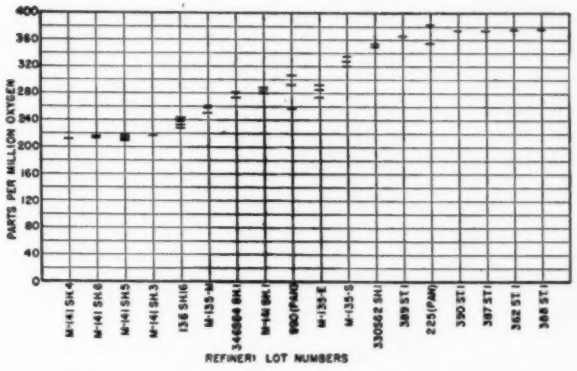


FIG. 3. Oxygen content of some Chilean coppers.

The general increase in oxygen content in the last five years can possibly be associated with a change in the composition of the ore in Chile. The ore has changed through the years from one of high purity oxide toward an increasing sulfide content. The sulfur is removed by the addition of oxygen to the melt. This treatment is followed by "poling" which introduces carbon and hydrogen to reduce the oxygen content to a desired value. The increased sulfur content of the ore could then result in a change in the oxygen content.

Sulfur

The increase in the sulfur content of the base ore suggests that this element may vary in refined Chilean copper. Experiments on cuprous-oxide rectifiers show that sulfur has marked effects on the characteristics. The results of the sulfur analyses are given in Fig. 5,

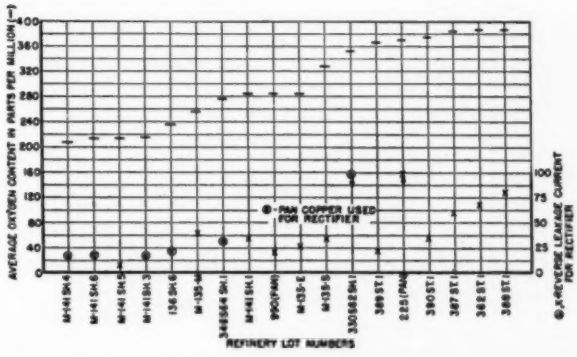


FIG. 4. Average oxygen content and cuprous-oxide rectifier reverse leakage current for some Chilean coppers.

[2] W. M. Hickam, Ind. Eng. Chem. (to be published).
[3] M. B. Beaver and C. F. Floe, Am. Inst. Mining Met. Engrs. 166, 128 (1946).

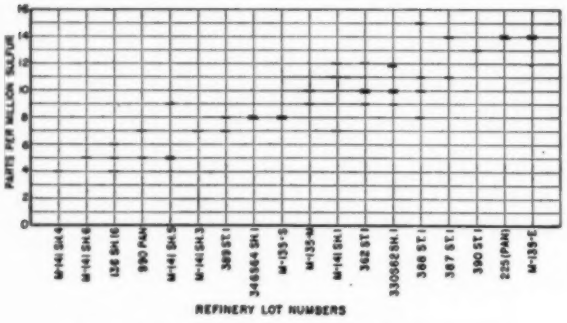


FIG. 5. Sulfur content of some Chilean coppers.

where the data has been arranged to yield increasing average sulfur content in going from left to right. Because of the spread in the analyses a large number of analyses were made in order to establish the average content of the various coppers. The spread in the analyses is believed due, in large part, to variations in the sulfur content of the samples and only a minor part due to errors of analysis. Variations are observed in sheets from M-141 and M-135 refinery lots. The large range in the sulfur content of the coppers, 4 to 15 parts per million, would suggest that variations could be expected in a single refining lot. In Fig. 6 is shown the average sulfur content and reverse leakage currents for various refining lots. In general the "old" coppers contain less sulfur than the "new" coppers. Figure 6 shows a correlation of rectifier reverse leakage current and sulfur content for most of the coppers. It is also noticed that, in general, the three factors—oxygen content, sulfur content, and reverse leakage current—follow the trend of increasing together. Since oxygen and sulfur are closely related chemically, it may not be too surprising that they both increase together.

In the processing of copper for cuprous-oxide rectifiers, various pretreating atmospheres have been used which yield marked changes in the rectifier. After establishing a general correlation between sulfur and oxygen content of untreated coppers and reverse leakage current, analyses were made to determine the effect of

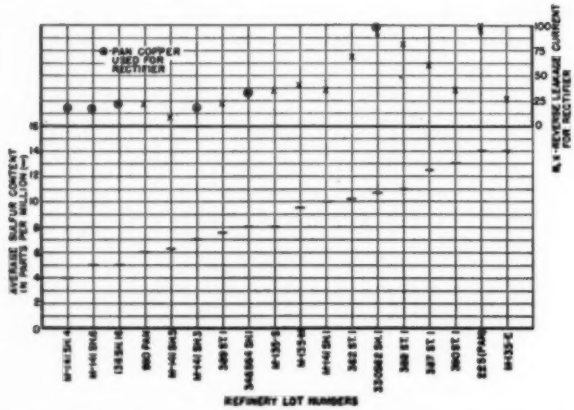

PAN COPPER USED FOR RECTIFIER
AVERAGE SULFUR CONTENT IN PARTS PER MILLION (+)
MA.=REVERSE LEAKAGE CURRENT FOR RECTIFIER
REFINERY LOT NUMBERS

FIG. 6. Average sulfur content and cuprous-oxide rectifier reverse leakage current for some Chilean coppers.

pretreatment in various atmospheres on sulfur and oxygen content. The results for the sulfur are given in Fig. 7. Runs 2, 3, 4, 5, and 6 yielded deoxidized copper. In general, samples of Chilean copper pretreated in hydrogen at 2-mm Hg pressure show a decrease in sulfur content. It should be noted that OFHC (non-Chilean) does not show this reduction. The fraction of the total sulfur removed is larger for Chilean copper samples of higher sulfur content. This may be associated with the general higher oxygen content. Samples pretreated in hydrogen at 1020°C and 2 mm Hg for four hours show no appreciable reduction in sulfur content below those treated for twenty minutes. The twenty-minute treatment is sufficient to deoxidize the copper completely and thus indicates the importance of oxygen in removing the sulfur. The failure of the hydrogen treatment to remove sulfur from the OFHC is in agreement with this. A third pretreatment in hydrogen lowering the temperature to 900°C was apparently not as efficient for removing the sulfur as at 1020°C. The pretreatment of disks in carbon

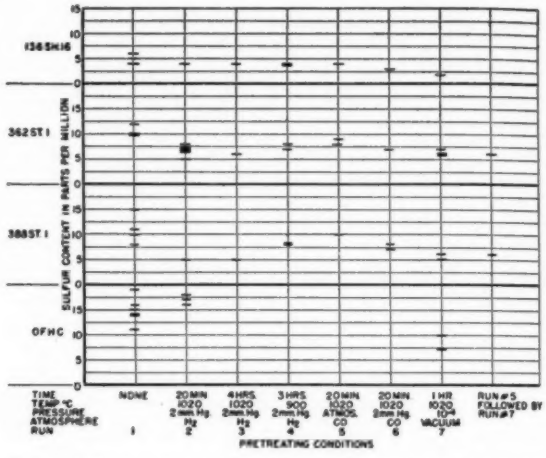


FIG. 7. Effect of pretreatment on sulfur content of copper.

monoxide at 1020°C and atmospheric pressure for twenty minutes showed, in general, no large decrease in sulfur content. A change in this treatment to a pressure of 2 mm-Hg did effect a lower sulfur content. The decrease in mean free path at the high pressure possibly prevented the escape of sulfur at the surface and the creation of a concentration gradient which is effective in transferring the sulfur to the surface. Vacuum treatment of the copper at 1020°C and 10^{-4} mm Hg for one hour does reduce the sulfur content but does not alter the oxygen content appreciably. These results would indicate that reduction in sulfur content is dependent upon both deoxidation and external pressure during heat treatment. The deoxidation of copper in vacuum, carbon monoxide, and hydrogen has been investigated by Ransley.[4]

In general, the low pressure hydrogen pretreatment has been found the most successful in bringing all

[4] C. E. Ransley, J. Inst. Metals **65**, 147 (1939).

coppers into the same classification as to rectification when processed as described. In making the rectifier the copper disks are pretreated first in the desired atmosphere and then processed in the described manner. The hydrogen pretreatment leaves the forward drop unaffected but does reduce the reverse leakage current appreciably. The pretreatment in vacuum and atmospheric pressure of carbon monoxide results primarily in increased forward drops as compared to those of untreated rectifiers. The hydrogen treatment results in the removal of oxygen and the reduction of sulfur content. This experiment emphasized the importance of these two elements in rectification.

In Fig. 8 are shown the sulfur contents for various treatments of half-thickness disks as compared to the original 0.050-in. thickness. The data indicate no large concentration of sulfur at the surface of the untreated disks as could be introduced in the rolling operation. In washers which were hydrogen treated and then etched to half thickness by removal of one-quarter thickness from each side of the disk, the sulfur content of this center half was found to be approximately the same as the untreated coppers. This shows that the low pressure hydrogen treatment leaves the copper nearly void of sulfur near the surface, but does not alter the concentration at a depth equal to or greater than one-fourth thickness. Etching to half-thickness followed by hydrogen or vacuum treatment did not result in sulfur content appreciably different from a similar treatment of the full thickness disks.

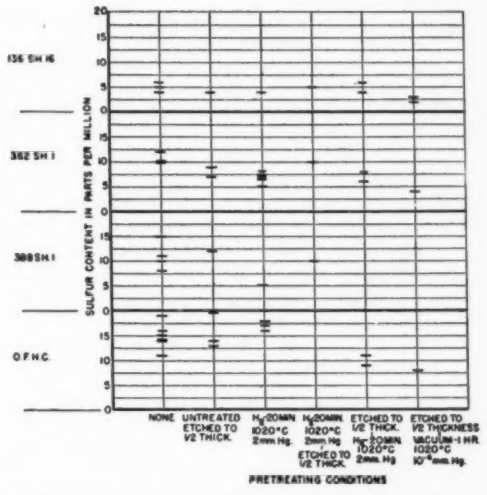

136 SH 16
362 SH 1
388 SH 1
OFHC
SULFUR CONTENT IN PARTS PER MILLION
NONE
UNTREATED
ETCHED TO V2 THICK.
H2 20MIN 1020°C 2mm Hg.
H2 20MIN 1020°C 2mm Hg ETCHED TO V2 THICK.
ETCHED TO V2 THICK. H2 20MIN 1020°C 2mm Hg
ETCHED TO V2 THICKNESS VACUUM-1 HR 1020°C 10⁻⁶mm Hg.
PRETREATING CONDITIONS

FIG. 8. Effect of pretreatment on sulfur content of copper.

III. CONCLUSIONS

(1) The carbon content of Chilean copper has not changed over the past ten years. It approximates one part per million and is in agreement with solubility data for a refining temperature of 1100°C.

(2) The oxygen content of Chilean copper has increased by approximately 50 percent in the past ten years. A general correlation has been found between the oxygen content and the reverse leakage current for cuprous-oxide rectifiers processed as described.

(3) The sulfur content of Chilean copper has increased by approximately 80 percent in the past ten years. A general correlation has been found between the sulfur content and the reverse leakage for cuprous-oxide rectifiers processed as described.

(4) Various pretreatments for copper have been shown to alter the sulfur and oxygen content. These pretreatments result in marked changes in the characteristics of the rectifier.

We wish to recognize the encouragement and guidance given us in these investigations by Dr. E. D. Wilson and the late Dr. R. C. Mason, of these Laboratories.

Enhanced Emission from Magnetron Cathodes*†

ROBERT L. JEPSEN AND MARCEL W. MULLER
Columbia Radiation Laboratory, Columbia University, New York, New York

(Received April 16, 1951)

Existing theories of the static magnetron do not predict violation of the Hull cut-off condition or cathode bombardment. Measurements made at anode voltages and magnetic fields much larger than those used by earlier investigators revealed violations of the Hull cut-off condition considerably more pronounced than those previously reported, and much greater amounts of cathode bombardment. When sufficient cathode bombardment takes place, emission may be enhanced (through secondary emission) by a factor of $\sim 10^3$ over the thermionic value. When enhanced emission occurs in magnetrons employing pure metal cathodes, it is found that maximum current limits exist for each magnetic field. These limits form a maximum current boundary on a voltage-current plot. The maximum current boundary depends on secondary emission properties of the cathode surface and on geometry. Experimental observations on enhanced emission, cathode bombardment, the maximum current boundary, and flow of anode current in violation of the Hull cut-off condition are combined into a self-consistent picture in static magnetrons employing pure metal cathodes. Although cathode bombardment and flow of anode current are still unexplained, their existence can account for the observed enhanced emission and maximum current boundary. In this it is unnecessary to endow the cathodes with anomalous thermionic or secondary emission properties.

The process causing cathode bombardment, enhanced emission, and flow of anode current in violation of the Hull cut-off condition is believed to be an electronic interaction which occurs almost entirely in the relatively dense region of space charge surrounding the cathode. The significant factors influencing the interaction appear to be applied electric and magnetic fields, space charge density, and thickness of the space charge sheath. The basic problem of electronic interaction in the static magnetron is partially defined. It is expected that when a realistic solution to this problem exists, including a correct account of cathode bombardment, the occurrence of enhanced emission would be automatically predicted.

I. INTRODUCTION

IN 1921 Hull[1] reported on attempts to verify experimentally the cut-off relation for a static magnetron. He observed that some current could be drawn at voltages below the cut-off voltage, though the amount was small in comparison with the available thermionic emission.

In 1943 Harvey[2] published results of careful measurements of departures from the Hull cut-off relation in a static magnetron. For cut-off voltages of a few hundred volts and thermionic emissions of ~ 0.01 ampere he found that ~ 20 percent of the available emission reached the anode for voltages ~ 80 percent of the cut-off voltage. For lower thermionic emissions the departures from the Hull condition were even greater.

In the early attempts to build magnetron oscillators it was found that a general tendency existed for the cathode to overheat. This was usually ascribed to some sort of bombardment of the cathode by charged particles, and various processes were suggested. In 1938 Linder[3] summarized some of these observations and reported that cathode bombardment occurred even in a static magnetron. He attributed the cathode bombardment, and the accompanying departure from the Hull cut-off relation, to an electron scattering process, but was unable to find a satisfactory mechanism.

The measurements reported by Hull, Linder, and Harvey were made on comparatively low voltage magnetrons ($V_a < 1000$ volts) with filamentary (small diameter) cathodes. Thermal dissipation problems would have made measurements very difficult at higher voltages.

Development of multi-cavity magnetrons was greatly accelerated during World War II. The necessity for getting usable sources of large amounts of rf power as rapidly as possible resulted, in some cases, in the construction, and even mass production, of magnetrons whose operation was only partially understood. The cathode emission requirements in modern short wavelength magnetrons are extremely severe; some magnetrons are now being operated with cathode current densities in excess of 50 amperes/cm^2. Surprisingly, conventional oxide-coated cathodes are capable of meeting these requirements;[4,5] similar cathodes employed in ordinary diodes would not give emission densities larger than about 0.5 ampere/cm^2 even under conditions of pulsed operation. If ϵ, the enhancement of emission, is defined as

$$\epsilon = \frac{\text{emission actually obtained}}{\text{thermionic emission available}},$$

it is seen that values of $\epsilon \sim 10^2$ are observed. This enhancement of emission was recognized, but no adequate explanation of the phenomenon was advanced. It was observed that of the order of 5 to 10 percent of the total input power was returned to the cathode as back bombardment.[6] It was also observed that secondary-

* Work supported jointly by the Signal Corps and the ONR under Signal Corps Contract W 36-039 sc-32003.
† Publication assisted by the Ernest Kempton Adams fund for physical research of Columbia University.

[1] A. W. Hull, Phys. Rev. **18**, 31 (1921).
[2] A. F. Harvey, *High Frequency Thermionic Tubes* (John Wiley and Sons, Inc., New York, 1943), pp. 83–113.
[3] E. G. Linder, Proc. Inst. Radio Engrs. **26**, 346 (1938).
[4] Fisk, Hagstrum, and Hartman, Bell System Tech. J. **25**, 257 (1946).
[5] G. B. Collins, *Microwave Magnetrons* (McGraw-Hill Book Company, Inc., New York, 1948), pp. 9, 23–24, 503–514.
[6] Reference 5, pp. 286, 525.

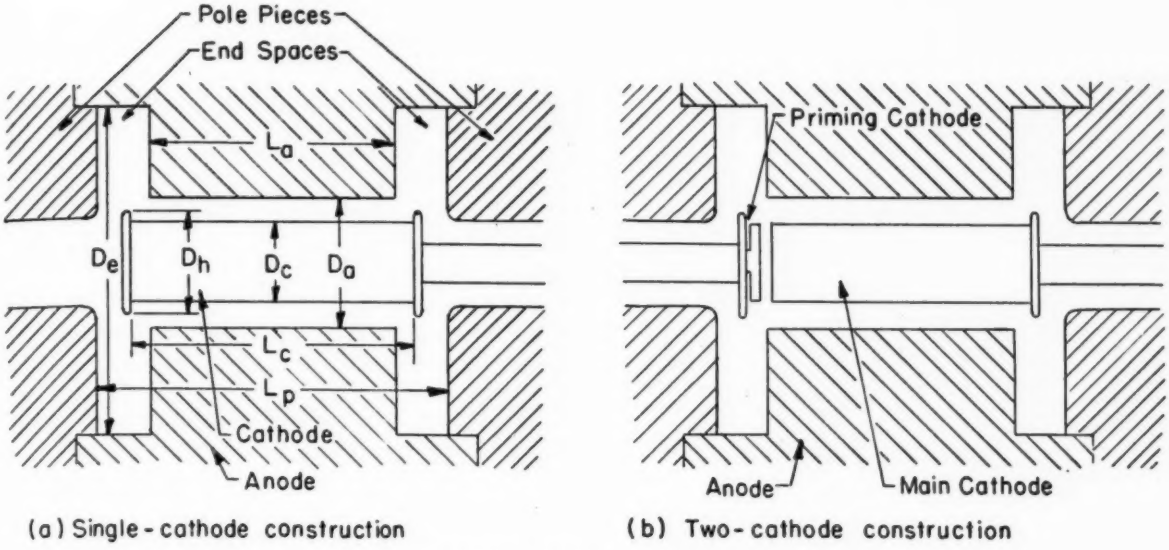


FIG. 1. Cross-sectional views of AX9's and AD9's showing significant geometrical parameters and single- and two-cathode constructions.

emission-type cathodes could be made to work when an auxiliary (priming) cathode was used to supply a small amount of thermionic emission.[7]

In view of the fact that phenomena connected with oxide-coated cathodes are still only incompletely understood, the present study of magnetrons and magnetic diodes employing pure metal cathodes was undertaken. It was hoped that the behavior of these relatively simple devices would shed some light on the process giving rise to enhanced emission. It became clear, as the work progressed, that the phenomena of enhanced emission, cathode back bombardment, and violation of the cut-off relations are intimately connected, and that the present data form a consistent extension into a higher voltage region (made accessible by pulsed techniques) of the results obtained by earlier investigators.

II. EXPERIMENTAL PROCEDURE

Three basic magnetron structures (see Fig. 1) were used for this study:

(1) Single-cathode 18-cavity rising sun X-band magnetron, designated as AX9-_D.

(2) Two-cathode version of AX9-_D, designated as AX9-_E.

(3) Two-cathode solid anode magnetron (static magnetron), designated as AD9-_E.

The significant geometrical parameters are (Fig. 1): $D_a=2r_a$=anode diameter, $D_c=2r_c$=cathode diameter, D_h=diameter of cathode hats, D_e=diameter of the end spaces, L_a=anode length, L_c=length of cathode between hats, and L_p=distance between the pole pieces (the magnetic gap).

In the AX9-_D, the molybdenum or tantalum cathode is heated indirectly by a heavy tungsten coil to a temperature of $\sim$2100°K to give a thermionic emission density of $\sim$0.05 ampere/cm^2.

In the AX9-_E and AD9-_E the two cathodes have the following functions:

(1) The priming cathode, located outside of the anode region, supplies the thermionic emission needed for current to build up at the beginning of each pulse. It is connected externally to the main cathode.

(2) The main cathode is located inside the anode region. It runs relatively cool, and supplies no significant thermionic emission. Anode voltage is applied between the cathodes and the anode. Upon electron bombardment the main cathode may emit secondary electrons. Molybdenum, platinum, nickel, copper, and carbon have been employed as main cathode materials.

Most of the measurements of interest were made at high voltages (up to 40 kilovolts). Because of power dissipation limitations in the magnetrons, pulsed voltages were used. (It has been ascertained that the phenomena of interest here exist under conditions of both pulsed and continuous operation.) Pulse widths

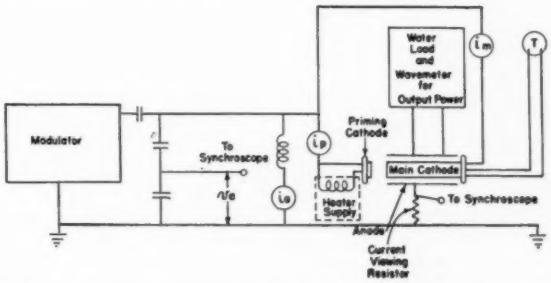


FIG. 2. Schematic representation of experimental setup used in most of the measurements described in this paper. v_a=fraction of anode voltage; voltage pulse height measured on synchroscope. i_a=average anode current. i_p=average priming cathode current. i_m=average main cathode current. T=temperature of thermocouple junction in main cathode.

[7] Reference 5, pp. 517–519.

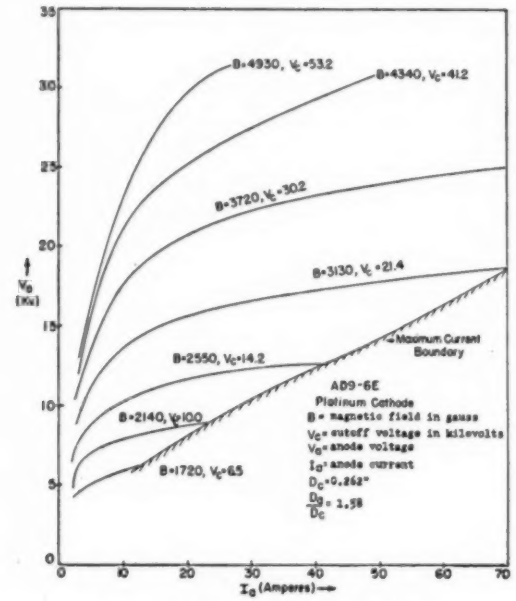


FIG. 3. $V_a - I_a$ curves for AD9-6E.

employed varied between 0.6 and 2.0 microseconds, and repetition rates between 120 and 960 cycles per second. Figure 2 shows schematically the experimental setup used.

Quantities of interest to be determined are: B = magnetic field, V_a = anode voltage, I_a = anode current, I_p = priming cathode current, I_m = main cathode current, P_b = back bombarding power, P_0 = output power, if any, and λ = wavelength of the output power.

Standard procedures were followed in measuring B, V_a, I_a, P_0, and λ. The measurement of I_p and I_m required the use of milliammeters, suitably by-passed, located in series with the priming and main cathodes.

The average back bombarding power, $\bar{P}_b$, can be conveniently determined by noting the temperature

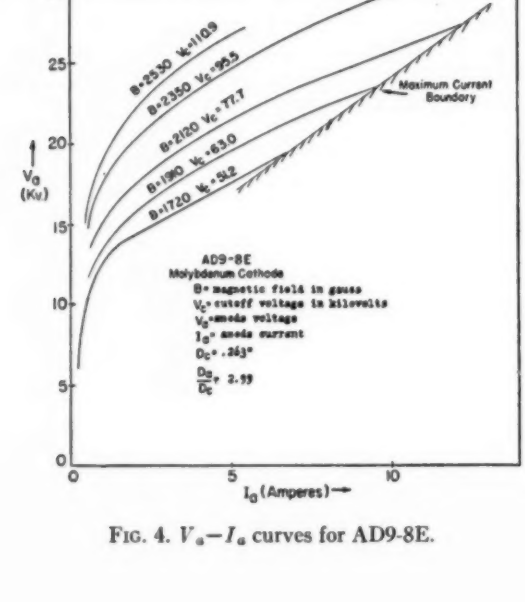


FIG. 4. $V_a - I_a$ curves for AD9-8E.

rise ΔT of a thermocouple located at the base of the main cathode. To calibrate the thermocouple, the anode voltage and magnetic field are turned off, and the main cathode is bombarded with a measured amount of power P_p by applying a DC voltage between the priming and main cathodes. P_p and $\bar{P}_b$ are not applied to exactly the same regions of the main cathode, therefore, $\bar{P}_b \neq P_p$. For the geometries and materials used, $\bar{P}_b < P_p$ for a given ΔT. On the basis of rough estimates it is believed that the errors involved in assuming $\bar{P}_b = P_p$ do not exceed 25 percent under the most unfavorable conditions of geometry, material, and temperature encountered here. For many comparison purposes these errors are hardly serious, as they tend to cancel out. If the need should arise for more accurate measurements of back bombarding power, in connection with a more refined theory, such measurements can be made, and corrections calculated. Random errors are estimated to be $\sim$3 percent for large average powers ($\sim$10 watts) and $\sim$25 percent for small average powers ($\sim$0.2 watt).

The magnetic field is measured in terms of the current in the coils of the electromagnet. Calibration is accomplished by measuring the field in a dummy which is magnetically identical with the magnetron. Magnetic field values, when given, are those measured at the geometric center. Random errors are estimated to be less than 1 percent.

Anode voltage measurements are estimated to be accurate to within $\sim$3 percent.

Peak values of anode current, input power, output power, and back bombarding power involve knowledge of the duty cycle. Inaccuracies in pulse width determination may result in errors in the duty cycle as large as 5 to 10 percent. Average values of anode current, output power, and back bombarding power, however, do not involve the duty cycle. Ratios of output power and back bombarding power to input power can therefore be computed without incurring this error.

III. RESULTS

A. Enhanced Emission from Pure Metal Cathodes

Earlier observations had shown that enhancement of emission (with $\epsilon \sim 10^2$) occurred in tubes employing oxide cathodes and secondary emission cathodes. It is found that enhanced emission (with ϵ as large as $\sim 10^3$) can also be obtained from pure metal cathodes. It is further found that enhancement of emission occurs in static magnetrons to about the same extent as in multicavity oscillating magnetrons.

B. The Performance Chart

Many of the properties of an oscillating magnetron are displayed on a plot called the performance chart. In this plot the magnetic field B is a parameter which takes on various discrete values. For each value of B a curve of V_a (anode voltage) *versus* I_a (anode current) is drawn. It is customary also to draw contours of constant output

power P_0 and contours of constant efficiency $(P_0/P_a) \times 100$ percent on the same chart. In the absence of output power, curves of V_a *versus* I_a can, of course, still be drawn.

According to the usual magnetron theories, anode current should not flow in a static magnetron before the Hull voltage is reached, nor in a multi-cavity oscillating magnetron until the Hartree voltage is reached. Substantial departures from these conditions are found, especially in the static magnetron. This is made clear by reference to Figs. 3 and 4, where the Hull voltage appropriate to a given magnetic field is indicated on each curve. It is seen that the departure from the Hull cut-off condition is relatively small at low voltages, but becomes greater at high voltages. It is also seen that the departure is more severe in static magnetrons having larger values of D_a/D_c.

C. The Maximum Current Boundary

When oxide cathodes are employed in magnetrons, operation is often possible out to very high currents. When pure metal cathodes are used, however, very

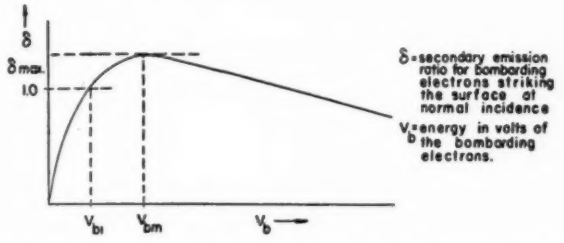


FIG. 5. Typical curve of secondary emission ratio *versus* incident electron energy (normal incidence).

definite current limits are found. For low enough values of magnetic field no enhancement of emission occurs. For larger magnetic fields and higher anode voltages it may be found that anode current can exceed the available thermionic current. The anode current continues to increase with voltage until a point is reached at which a further increase in voltage reduces the current abruptly to, approximately, the thermionic value. Corresponding to each magnetic field there exists a maximum current. The locus of these maximum current points is said to be the *maximum current boundary* (MCB). It is usually found that the MCB is very nearly a straight line, except in the low current region. If a linear extrapolation to zero current is made, it is found that the voltage axis is intersected near 4 kilovolts, depending on the pure metal used and on some of the geometrical factors involved. It is also found that the position of the MCB is virtually unaffected by the available thermionic emission (except in the very low current region), unless the thermionic emission falls below the critical value needed for buildup to occur. The position of the MCB is found to be extremely sensitive, however, to the secondary emission properties of the main cathode surface.

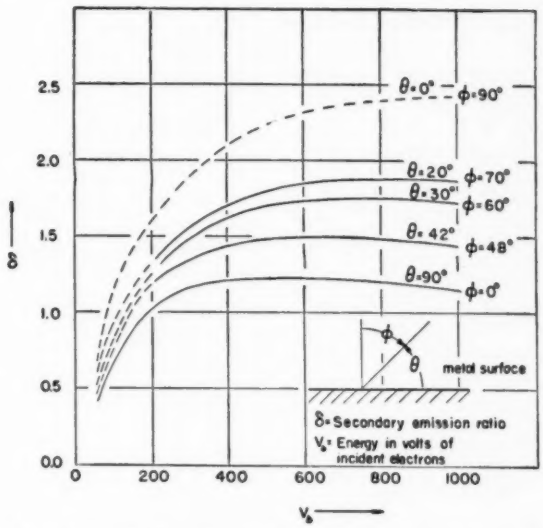


FIG. 6. Secondary emission ratio of nickel as a function of electron energy for various angles of incidence (from H. Bruining, Physica **5**, 909 (1938)).

Most substances exhibit secondary emission properties of the form shown in Figs. 5 and 6 and in Table I. A convenient representative measure of secondary emission properties is δ_{max}, defined in Fig. 5. Dependence of the position of the MCB on δ_{max} was studied using oscillating AX9's. Figure 7 shows MCB's corresponding to four different pure metals—tantalum, molybdenum, platinum, and nickel—and to carbon. It is seen that the slope of the MCB curve is smaller for larger δ_{max}. No enhanced emission at all was observed with carbon, for which $\delta_{max} \simeq 0.9$. The largest enhanced emission was observed with platinum, for which $\delta_{max} \simeq 1.7$–1.8.‡[8]

TABLE I. Secondary emission properties of several pure metals (compiled from H. Bruining, *Die Sekundär-Elektronen-Emission fester Körper* (Verlag. Julius Springer, Berlin, 1942) (Edwards Brothers, Inc., Ann Arbor, Michigan, 1944), pp. 21–27). For explanation of symbols, see Fig. 5.

Metal	δ_{max}	V_{bm}	V_{b1}
Molybdenum	1.2–1.3	400 volts	150 volts
Tantalum	1.2–1.3	600–700	300
Tungsten	1.2–1.4	500–700	200–300
Niobium	1.2	350	200
Platinum	1.7–1.8	750	150
Nickel	1.3–1.4	500	200
Copper	1.2–1.3	300–500	100–200
Silver	1.5	800	200
Gold	1.2–1.4	400–1100	200
Titanium	0.9	300	—

‡ The present observations on the dependence of the MCB on secondary emission properties form a consistent extension of some studies by McNall, Steele, and Shackelford (see reference 8) on the influence of secondary emission ratio properties of the cathode surface on magnetron operation.

[8] McNall, Steele, and Shackelford, "Secondary emission cathodes for magnetrons," Final Report on Contract OEMsr-262, Sub-Contract DIC 182008, Research Report BL-R-929-7G-1, November 9, 1945.

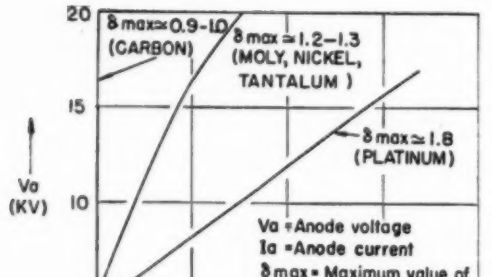

δ max≃0.9-1.0 (CARBON)
δ max ≃1.2-1.3 (MOLY, NICKEL, TANTALUM)
δ max≃1.8 (PLATINUM)
Va (KV)
Va =Anode voltage
Ia =Anode current
δ max = Maximum value of secondary emission ratio for normal incidence.
Ia (amps)

FIG. 7. Maximum current boundaries for oscillating AX9's with main cathodes having different secondary emission properties.

For a given cathode material the position of the MCB is fairly constant in time. Arcing which releases a small amount of gas sometimes causes the MCB to move temporarily, usually to higher currents. This is attributed to a reaction at the cathode surface which changes the secondary emission properties slightly. It is believed that under conditions of good vacuum and arc-free operation the secondary emission properties of the cathode surface approach the stable values appropriate to the pure metal involved.

On the basis of the preceding information it is known how the MCB depends on δ_{max}. The dependence of δ_{max} on angle of incidence is shown in Fig. 6. To see if electrons strike the cathode within some preferred angle range, a cathode surface was machined in an asymmetrical fashion, as indicated in Fig. 8. If the electrons come in within a preferred angle range, there will clearly be a dependence of effective angle of incidence—and therefore of δ_{max} and the MCB—on magnetic field direction. An example of the observed dependence of the MCB on magnetic field direction is shown in Fig. 9. Thus far all that can be reliably concluded from experiments using asymmetrical cathodes is that a significant number of the back bombarding electrons strike the cathode within some range of angles of incidence rather than normally or at random angles.

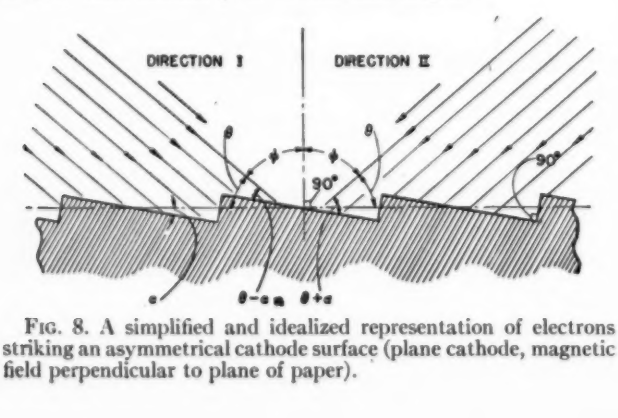


FIG. 8. A simplified and idealized representation of electrons striking an asymmetrical cathode surface (plane cathode, magnetic field perpendicular to plane of paper).

A comparison of the MCB's for oscillating multicavity magnetrons and static magnetrons is shown in Fig. 10. Except for resonators and output the geometry of the two magnetrons is similar. It is seen that the positions of the MCB's are not greatly different. This suggests that the process which causes enhanced emission is largely independent of the anode.

In a further effort to find to what extent conditions at the anode affect the MCB, a carbon anode tube was built. End spaces were considerably reduced in size, and were also made of carbon. The high resistivity of carbon (relative to the copper normally used) should reduce the Q of a circuit involving the anode and end spaces by a factor of $\sim$30. It was found that enhanced emission occurred as usual, and that the position of the MCB was not noticeably affected.

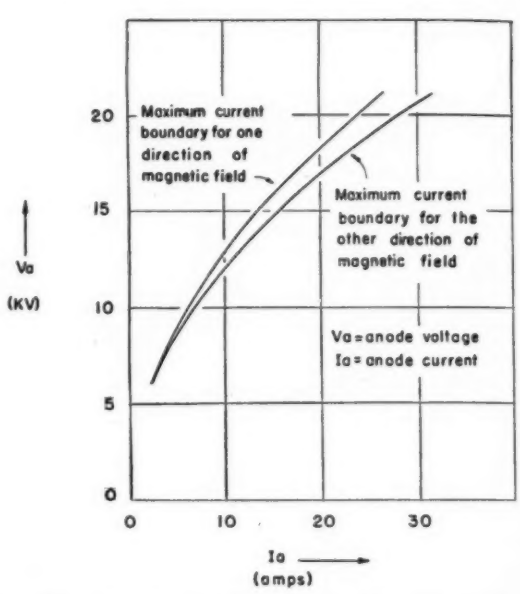

Maximum current boundary for one direction of magnetic field
Maximum current boundary for the other direction of magnetic field
Va=anode voltage
Ia=anode current
Va (KV)
Ia (amps)

FIG. 9. Maximum current boundaries for the two directions of magnetic field. Data from AX9-27.1E.

The dependence of the MCB on geometry is indicated in Fig. 11, with values of the geometric parameters given in the accompanying table. AD9-1E is taken as a tube of standard geometry. AD9-3E differs in that it has a half-length anode and its MCB occurs at currents which are approximately one-half as great as those of the standard tube. This suggests that the process responsible for the enhanced emission does not depend in a critical way on anode length. AD9-8E, AD9-9E, and AD9-11E involve different combinations of anode and cathode diameters. It is noted that increasing the anode diameter, leaving the cathode diameter fixed, results in an increase in the slope of the MCB curve. In the case of AD9-9E and AD9-11E, where the anode and cathode diameters are both decreased by a factor of 2, with the ratio D_a/D_c held constant, the slope of the MCB decreases by approximately a factor of 2. Thus, in view of the dependence of the MCB on anode length, and

extrapolating somewhat boldly, it appears that changing *all linear dimensions* by a constant factor will leave the position of the MCB substantially unchanged.§

D. Cathode Bombardment

In the interest of finding out what goes on inside a magnetron it would be useful to know the number of electrons returning to the cathode per unit time and their energy distribution. Such detail is difficult to determine, but useful information concerning cathode bombardment is contained in the amount of back bombarding power incident upon the cathode. This information, when combined with enhanced emission data, does give some idea of an equivalent back bombarding voltage and current.

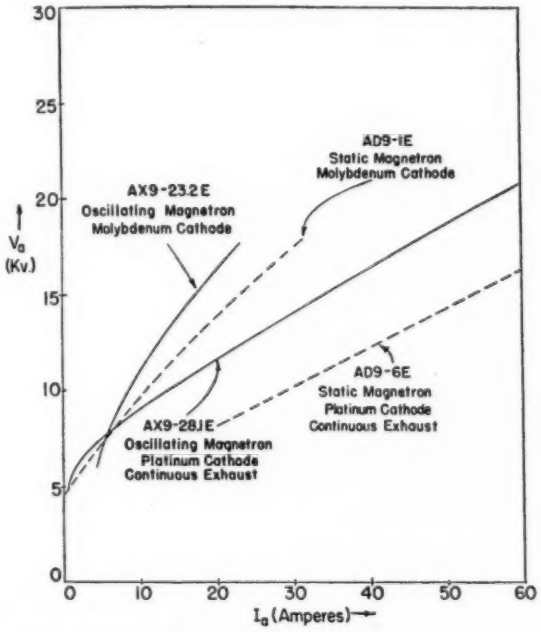
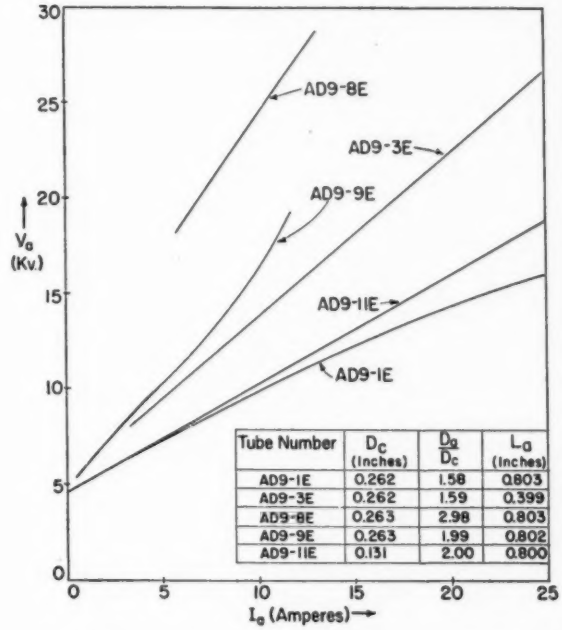


FIG. 10. Comparison of maximum current boundaries for static and oscillating magnetrons of similar geometries.

On the basis of earlier measurements by others, it had been generally accepted that ~5 to 10 percent of the total input power (P_a) goes into back bombardment in oscillating magnetrons.[6] Apparently, most of such measurements were confined to the relatively high current regions of the performance chart. It turns out that the present method for measuring back bombarding power is sensitive down to very low currents. Back bombarding power is easily measured at currents which are smaller by a factor of ~30 than those at the oscillation threshold. If cathode bombardment is to be ascribed to interaction of out-of-phase electrons with

§ This result is also in agreement with some earlier observations made on oscillating multi-cavity magnetrons operating at wavelengths of 3.2, 1.2–1.3, and 0.62 cm. These tubes differed primarily in that their linear dimensions had been changed by a constant factor (wavelength scaling). It was observed that the MCB's were practically the same for all of these tubes.



Tube Number	D_c (Inches)	$\frac{D_a}{D_c}$	L_a (Inches)
AD9-1E	0.262	1.58	0.803
AD9-3E	0.262	1.59	0.399
AD9-8E	0.263	2.98	0.803
AD9-9E	0.263	1.99	0.802
AD9-11E	0.131	2.00	0.800

FIG. 11. Maximum current boundaries corresponding to some different geometries. All main cathodes are molybdenum.

the rf field of an oscillating magnetron, then it is to be expected that where there is no rf, there should be no back bombardment. In a simple way it might be supposed that P_b should be approximately proportional to P_0. Thus if P_0/P_a as a function of I_a (for a given B) has the form indicated by curve A of Fig. 12, then it might be expected that P_b/P_a would have the form of curve B. But measurements show that the actual form of the dependence is that indicated by curve C. It is further found that the same sort of cathode bombardment occurs even in the static magnetron. The occurrence of back bombardment in the absence of rf of anything like the conventional sort has led to a re-examination of a number of earlier notions regarding both static and multi-cavity oscillating magnetrons.

A comparison of back bombarding power in static and in oscillating multi-cavity magnetrons of similar geometries is given in Figs. 13 and 14. It is seen that the shapes of the curves of P_b/P_a *versus* I_a are very similar

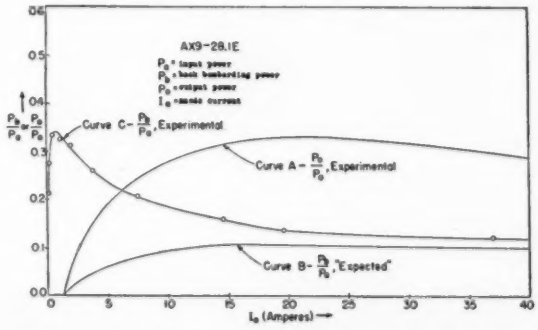


FIG. 12. Power output and back bombarding power for AX9-28.1E at a magnetic field of 3730 gauss.

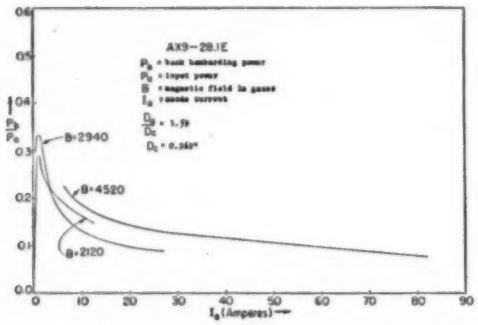


FIG. 13. Back bombarding power curves for AX9-28.1E (platinum main cathode).

in the two cases. The significant difference appears to be that P_b/P_a is somewhat higher for the static magnetron.

Neither the anode material nor the cathode material seem to affect the cathode bombardment, except that measurements can be made to higher currents in a tube with a main cathode having a higher δ_{max}. Changing the ratio D_a/D_c has a large effect, P_b/P_a decreasing rapidly as D_a/D_c is increased.

E. Division of Current Between Priming and Main Cathodes

In some cases the two cathode currents were measured separately. A typical result is shown in Fig. 15. As the anode voltage V_a is raised, the main cathode current I_m is at first increasingly negative (i.e., there is a net flow of electrons *to* the main cathode); it goes through a negative maximum, and then starts to increase, through zero, until the MCB is reached. While I_m is still negative, P_b starts from zero and increases rapidly with V_a.

The interpretation of the main cathode current reversal is the following: for very small values of P_b the energies of the back bombarding electrons are rather low, whence the effective secondary emission ratio of the cathode surface is small. Hence there is a net flow of electrons *to* the main cathode. As P_b increases, there is an accompanying increase in the energies of the back bombarding electrons. When the energies are such that the effective secondary emission ratio is unity, as many electrons leave the cathode surface as strike it, and the main cathode current is zero. Further increase in

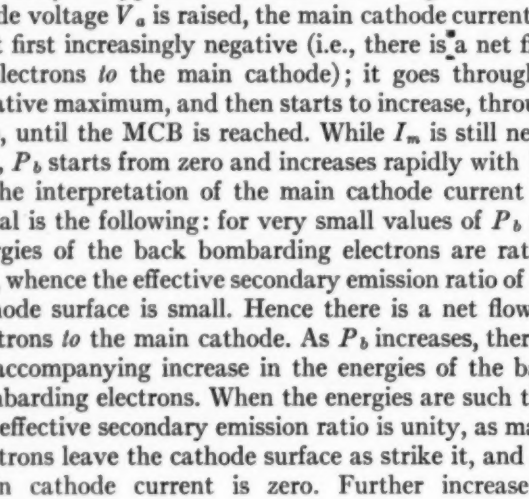


FIG. 14. Back bombarding power curves for AD9-4E (molybdenum main cathode).

energies of the back bombarding electrons results in an effective secondary emission ratio greater than unity, so that there is a net flow of electrons *from* the main cathode.

Consistent with the foregoing interpretation is the observation that when the main cathode was made of carbon no such current reversal occurred, the main cathode current being always negative. This is not surprising, since for carbon $\delta_{max} \simeq 0.9$.

For low enough values of magnetic field I_m does not reverse in tubes containing molybdenum or platinum cathodes, even though some back bombardment occurs. The lack of emission enhancement prevents P_b from becoming large.

IV. INTERPRETATION

A. An Explanation of Enhanced Emission

Some experimental observations on cathode bombardment, enhanced emission, current division, the

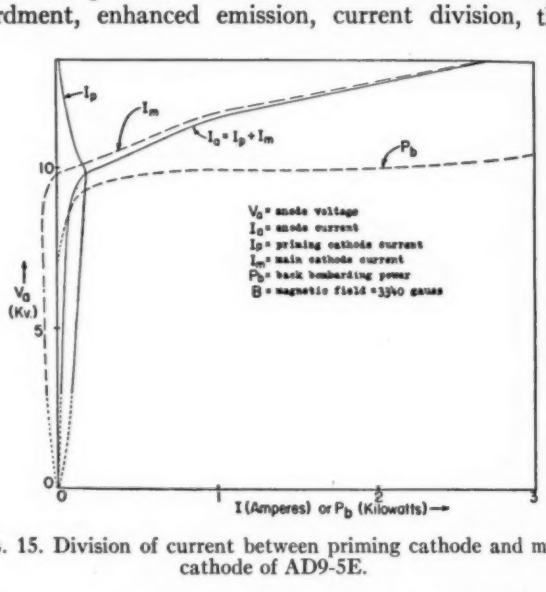


FIG. 15. Division of current between priming cathode and main cathode of AD9-5E.

MCB (maximum current boundary), and the flow of anode current have been described in the preceding section. It is possible to formulate a rather simple self-consistent picture which gives a qualitative explanation of enhanced emission. For this formulation the following basic assumptions are made:

(1) At any voltage and current point under consideration a condition of steady operation has already been achieved (i.e., the buildup process is ignored).

(2) The only significant sources of emission are thermionic and secondary. For pure metals it will suffice to consider only *stable* emission properties—no anomalously low work functions or high secondary emission ratios.

(3) Secondary emission occurs under the stimulus of the empirically observed cathode bombardment.

It would be desirable to know the distribution in energy and angle of the electrons striking the cathode. In the absence of this information, and in the interest of

simplifying the present explanation, it will be supposed that the state of cathode bombardment is characterized by the following: there is an electron current I_b striking the cathode within a narrow range of energy and angle. As an approximation, assume that these electrons all arrive at the cathode with the same energy, V_b. In addition, there may be a current I_0 of electrons emitted in excess of the requirements for operation which are returned to the cathode at low energy. It is supposed that these do not contribute significantly to the back bombarding power P_b, whence $P_b \simeq I_b V_b$. No inquiry will be made into the process responsible for the bombardment. Leakage current to the pole pieces is also ignored.

In view of these approximations (narrow ranges of energy and angle of incidence) it is convenient to define an effective secondary emission ratio, δ_e, which is really a properly weighted average over the angle of incidence. While δ_e cannot be actually calculated from the available information, it is reasonable that it should be similar in form to the measured secondary emission ratios at some definite angle of incidence. Thus δ_e depends on the substance involved, on V_b, and on the distribution over angle of incidence. Figure 19 shows an estimated curve of δ_e against V_b. If δ_e is sufficiently small so that there is little excess emission, then I_0 is small. If δ_e is larger so that there is more excess emission, then I_0 is larger. It is to be noted that the prediction of the existence of a MCB depends only on the general shape of the curve of Fig. 19. By subsequently using the position of the MCB as an experimental datum and working back, the curve can be fixed more accurately.

The precise dependence of V_b, or V_b/V_a, on anode voltage V_a or anode current I_a is not known. However, for all I_a for which $P_b=0$ it is expected that $V_b=0$, and for the I_a for which the main cathode current I_m goes through zero and changes sign it is expected that V_b is such that δ_e is approximately unity (except for the effect of I_0). V_b needs to be $\sim$100–300 volts for $I_m=0$, depending on the pure metal used and the effective angle of incidence.

Figure 16 shows semi-empirical curves of P_b/P_a and V_a *versus* I_a for several values of magnetic field for a static magnetron. Figure 17 shows estimated curves of V_b and V_b/V_a *versus* I_a. For the smaller values of I_a the forms of the curves are fairly unambiguous, in view of the above discussion. It will now be shown using the empirical curves of P_b/P_a *versus* I_a, the estimated curves of V_b/V_a *versus* I_a, and an estimated curve of δ_e *versus* V_b,‖ that a MCB will be predicted which will have, qualitatively, the experimentally observed shape and dependence on secondary emission properties. This, of course, will not prove uniqueness of the explanation of enhanced emission, but will add to its plausibility. A

‖ The curve of δ_e *versus* V_b will depend on angles of incidence, which probably have a magnetic field dependence. The latter dependence is not known, and is ignored here. Therefore the same curve will be used for all magnetic fields.

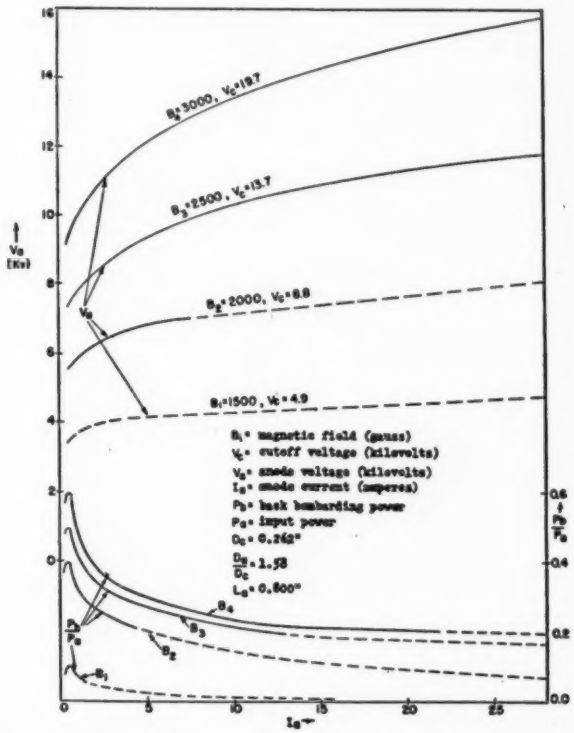


FIG. 16. Semi-empirical curves of P_b/P_a and V_a *versus* I_a for a representative AD9-_E static magnetron.

major point in this treatment is to show that the explanation of enhanced emission proposed here is at least energetically feasible. After showing how cathode bombardment, secondary emission properties, and the MCB are connected, it would then be possible to use the curves of P_b/P_a *versus* I_a and the MCB as experimental quantities from which to determine ranges of values for

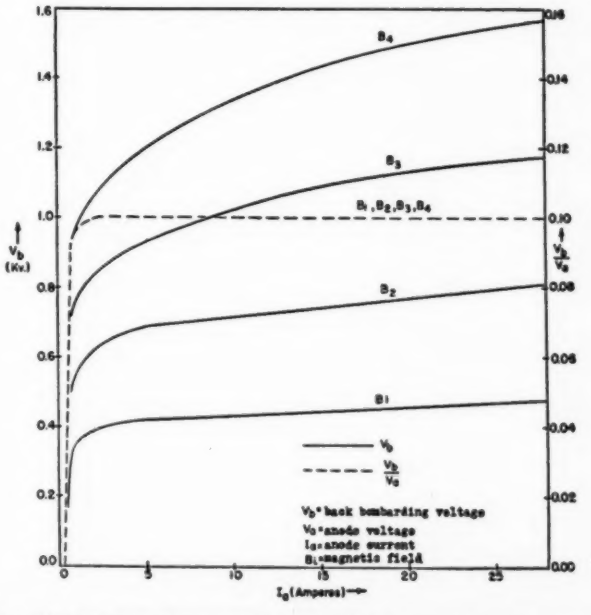


FIG. 17. Estimated curves of V_b and V_b/V_a *versus* I_a for a representative AD9-_E static magnetron.

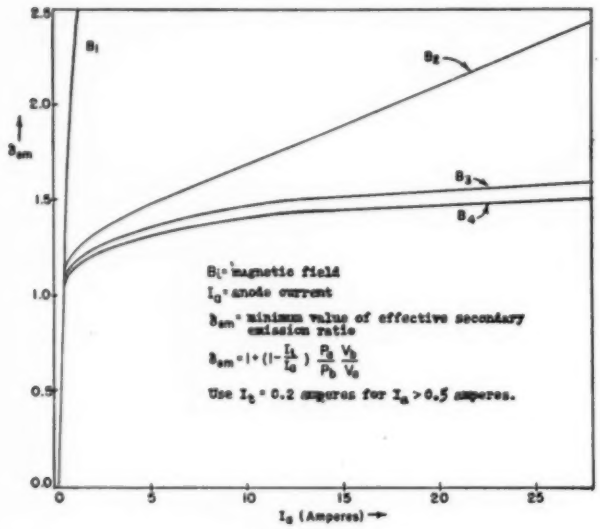


FIG. 18. δ_{em} *versus* I_a for a representative AD9-_E static magnetron.

V_b and angles of incidence. This would be done by adjusting the curves of δ_e *versus* V_b and V_b *versus* I_a to give "best fits." The results of such a procedure must, of course, not be taken too seriously, as they are based on the approximation of representing the actual energy distribution of the back bombarding electrons by a single energy V_b.

To facilitate further discussion the following definitions are given: I_t = temperature-limited value of the thermionic emission, I_b = back bombarding current (composed of electrons having energy V_b), I_a = anode current, δ_e = an effective secondary emission ratio; depends on the substance involved, on V_b, and on the effective angle of incidence, V_a = anode voltage, V_b = an effective value of back bombarding electron energy, $P_a = I_a V_a$ = input power, $P_b = I_b V_b$ = back bombarding power, and B = magnetic field.

For this discussion it is convenient to use B as a parameter and I_a as the independent variable. Thus for a given B an operating point is defined by I_a.

It is clear that the total available emission—thermionic plus secondary—must equal or exceed the required emission.

Thus,

$$I_t + \delta_e I_b \geqslant I_a + I_b.$$

Solving for δ_e as a function of I_a (I_b depends on I_a),

$$\delta_e \leqslant 1 + (I_a - I_t)/I_b.$$

Now define δ_{em} as

$$\delta_{em} = 1 + (I_a - I_t)/I_b.$$

If operation is to occur at a given I_a through a combination of thermionic and secondary emission, it is necessary—for current conservation—that $\delta_e \geqslant \delta_{em}$. If $\delta_e < \delta_{em}$, inadequate emission will exist for operation at this point. On this basis it might be expected that curves of $\delta_e = \delta_{em}$ will exist on the $V_a - I_a$ performance chart. On one side of the curve, or boundary, $\delta_e < \delta_{em}$ and operation is not allowed. On the other side, $\delta_e > \delta_{em}$, and operation will occur, provided there is not some further restriction (such as might be imposed by the electron interaction process which causes the cathode bombardment). Since little is known about these further restrictions, they will be ignored here, and it will be assumed that the boundary between regions of no operation and possible operation occurs at $\delta_e = \delta_{em}$.

Location of regions of possible operation in the $V_a - I_a$ performance chart will depend on how δ_e and δ_{em} vary with I_a for various magnetic fields. It turns out that, for

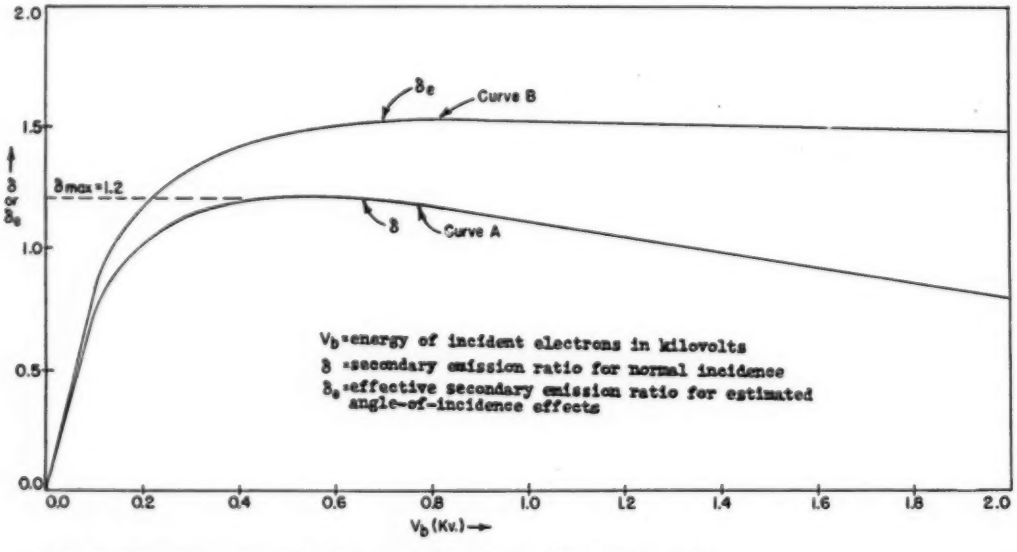

δe
Curve B
δmax = 1.2
δ or δe
δ
Curve A
Vb = energy of incident electrons in kilovolts
δ = secondary emission ratio for normal incidence
δe = effective secondary emission ratio for estimated angle-of-incidence effects
Vb (Kv.)

FIG. 19. Curve A: δ *versus* V_b for a typical pure metal having $\delta_{max} \simeq 1.2$.

Curve B: δ_e *versus* V_b upon taking account of estimated angle-of-incidence effects in a representative AD9-_E static magnetron.

suitable cathode material and judicious estimates of V_b and angles of incidence, the curve of $\delta_e=\delta_{em}$ occurs about where the experimentally observed maximum current boundary is found, and that for a given magnetic field $\delta_e>\delta_{em}$ for $I_a<I_a'$, and $\delta_e<\delta_{em}$ for $I_a>I_a'$, where $I_a=I_a'$ for $\delta_e=\delta_{em}$. Thus the curve of $\delta_e=\delta_{em}$ turns out to be a maximum current boundary.

This will now be gone through in detail for a specific static magnetron. It is convenient to write δ_{em}, using $P_a=I_aV_a$ and $P_b=I_bV_b$, as

$$\delta_{em}=1+(1-I_t/I_a)(P_a/P_b)(V_b/V_a).$$

From the experimental curves of P_b/P_a *versus* I_a, the experimental value of I_t, and the estimated curves of V_b/V_a *versus* I_a, curves (corresponding to various magnetic fields) of δ_{em} *versus* I_a may be constructed. These are shown in Fig. 18. For all I_a for which $\delta_e>\delta_{em}$ operation is allowed. For the I_a for which $\delta_e<\delta_{em}$ operation cannot occur.

The next step is to determine δ_e as a function of I_a. This depends on the pure metal being considered, an

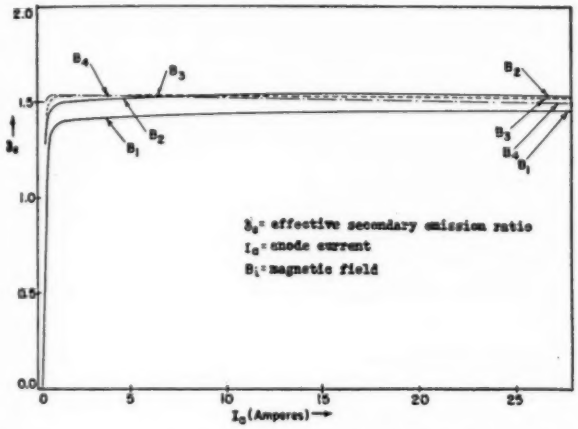


FIG. 20. δ_e *versus* I_a for a representative AD9-_E static magnetron.

estimate of the angle of incidence effects, and use of the estimated curves of V_b/V_a *versus* I_a. In Fig. 19, curve A shows δ *versus* V_b for a pure metal for which $\delta_{max}=1.2$, and curve B shows a modification which takes account of estimated angle of incidence effects (δ_e *versus* V_b). This is combined with the estimated dependence of V_b on I_a (Fig. 17) to give δ_e as a function of I_a for various magnetic fields, as shown in Fig. 20.

A superposition of the curves of Figs. 18 and 20 (Fig. 21) shows that, for $B=B_i$, operation is possible for $I_a<I_{ai}$ and not allowed for $I_a>I_{ai}$. (For operation, need $\delta_e>\delta_{em}$.) Thus maximum current points occur. It is a simple matter to transfer these maximum current points to the V_a-I_a performance chart (Fig. 22). It is observed that the resulting maximum current boundary (MCB) has the experimentally observed form, and occurs at approximately the right values of V_a and I_a. It should be remarked that the qualitative form of the MCB is fairly insensitive to the estimated curves of

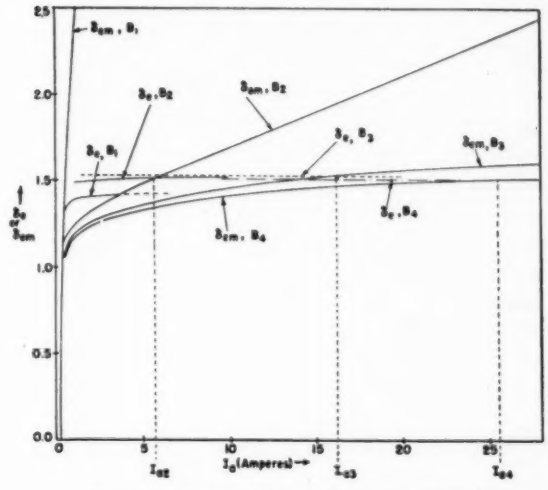


FIG. 21. Superposition of Figs. 18 and 20.

V_b/V_a *versus* I_a. However, the better the estimates of V_b/V_a *versus* I_a and δ_e *versus* V_b, the better the expected quantitative agreement.

From a brief reconsideration of the foregoing arguments, together with reference to the figures involved, it is clear that if the pure metal has a larger δ_{max}, the corresponding MCB will occur at higher currents, in agreement with observations on the correlation of position of the MCB with δ_{max} (Fig. 7).

Thus, it appears that the experimental observations on cathode bombardment, enhanced emission, current division, the maximum current boundary, and the flow of anode current can be integrated into a self-consistent

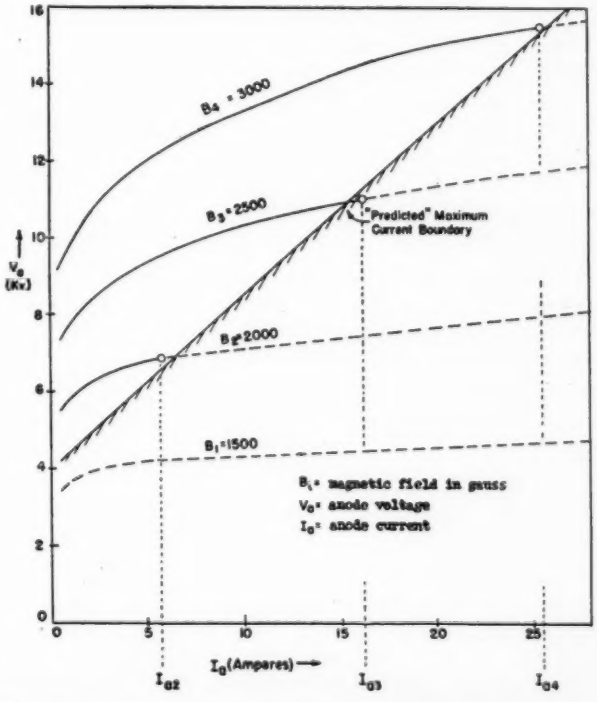


FIG. 22. V_a-I_a curves and predicted maximum current boundray for a representative AD9-_E static magnetron.

pattern. The cathode bombardment and flow of anode current are still unexplained, but their existence can account for the observed enhanced emission and the maximum current boundary. The quantitative agreement between the predicted and observed maximum current boundaries shows, essentially on energy conservation grounds, that enhanced emission can be accounted for in a straight-forward manner without endowing the pure metal cathodes with any anomalous emission properties. This by no means demonstrates uniqueness of the explanation, as it is undoubtedly possible to account for enhanced emission in other ways. Thus far, however, no simpler alternative explanation has been proposed.

B. Electron Interaction in the Static Magnetron

It now appears probable that cathode bombardment, enhanced emission, and flow of anode current in violation of the Hull cut-off condition are all manifestations of the same basic phenomenon. Probably a further manifestation is excess noise, which has not been investigated here to any significant extent. It should be remarked that the present observations on cathode bombardment and violation of the Hull cut-off condition are not inconsistent with earlier observations by other workers.[1–3] Their measurements were made at relatively low magnetic fields, voltages, and currents under dc conditions. The present measurements were made at comparatively high magnetic fields, voltages, and currents under pulsed conditions. Cathode bombardment and violation of the Hull cut-off condition become more pronounced with increasing magnetic fields and voltages. When cathode bombardment becomes sufficiently severe, enhanced emission can occur. The reason earlier workers did not observe enhanced emission is that their measurements were made at values of magnetic field and voltage which were too low for sufficiently energetic cathode bombardment.

On the basis of the observations made to date it appears that the phenomenon which causes cathode bombardment, etc., is characterized by the following: the electrons are involved in an interaction which depends on the cathode as an emitter, and on the anode as a source of applied potential. It is largely independent of length of anode, of end spaces, and losses in cavities that might form part of a resonant circuit. There is no appreciable dependence on gas pressure P for $P<\sim 10^{-3}$ mm of Hg.

For a given magnetic field and geometry it is found that at low anode voltages no readily observable cathode bombardment occurs. As the anode voltage is raised, cathode bombardment begins rather abruptly and increases rapidly. (See, e.g., Fig. 15.) In the voltage range of negligible cathode bombardment it may, perhaps, be plausible to say that something like a single stream steady state[9] exists. Once cathode bombardment has begun, and anode current has started to flow, the single stream steady state is clearly inapplicable. It may, however, be that the space charge density, the charge cloud radius, and the potential distribution are still described with reasonable accuracy by the single stream steady state. A comparison of magnetrons having the same applied magnetic fields, similar cathodes, but different anode diameters shows that similar states of cathode bombardment (and hence, presumably, similar states of electronic interaction) occur for equal space charge cloud radii, as calculated from single-stream theory.[10]

To sum up briefly, it appears that the interaction process in which the electrons in a static magnetron are involved occurs almost entirely among electrons in the relatively dense region of space charge surrounding the cathode, and that the significant factors influencing the interaction are the applied electric field, the applied magnetic field, the space charge density, and the thickness of the space charge sheath. To a first approximation the outer boundary of the space charge sheath may be taken as that predicted by single stream steady state theory.

It is a major objective of present magnetron research to account in a satisfactory way for the electronic interactions in the so-called static magnetron. From a theoretical point of view there are basically only two possible avenues of approach to the problem of electron interaction in a high density space charge. One is statistical, and is analogous to the methods used in kinetic theory or in statistical mechanics. The other is essentially a hydrodynamical approach.

Two crude close-collision calculations were done in which estimates were obtained of the number of electrons that might be expected to strike the cathode with energies greater than ~ 50 volts. The calculations were based on cycloidal orbits and on random orbits, and an average space charge density indicated by single stream theory was used. (If a single-stream steady state existed, no collisions at all would occur, since the electrons move in concentric circular paths.) The numbers thus obtained of significantly energetic electrons striking the cathode were smaller than observed by a factor of something like 10^6. As a consequence of these results there may be some justification for assuming that close collisions are not a major cause of the observed interaction, although the argument is by no means conclusive. A more refined treatment might be attempted which takes as its basis Boltzmann's equation of change of the distribution function for electrons in the presence of electric and magnetic fields.[11] This does not appear to be a simple problem.

In view of the difficulties connected with a statistical calculation it is not surprising that most successful investigations of space charge problems have used a

[9] L. Brillouin, Phys. Rev. **60**, 385 (1941).

[10] CRL Quarterly Report, December 31, 1950.

[11] S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases* (The University Press, Cambridge, England, 1939).

hydrodynamical approach. In these investigations the discrete nature of the electrons is ignored and the space charge is treated as a continuous medium. Thus far, however, no hydrodynamical theory has been set forth which describes the electron interaction in a static magnetron. It has been suggested by a number of workers that the interaction might be understood in terms of double-stream amplifier action,[12,13] or in terms of Bunemann space charge instability.[14] In connection with a theory on excess noise in preoscillating multicavity magnetrons, Twiss[15] took account of emission velocities and obtained a sort of double stream state for which space charge density and geometry are about the same as those given by single stream theory. In this double stream state there is electron motion in both the positive and negative radial directions, and radial double-stream amplification can take place when the radial current densities are favorable. Whether or not a suitable extension or modification of this theory can lead to a correct account of the electron interactions, with adequate prediction of cathode bombardment and anode current, is not yet known.

V. CONCLUSIONS

A. Summary

Experimental observations on enhanced emission, cathode bombardment, division of current between priming and main cathodes, the maximum current boundary, and flow of anode current in violation of the Hull cut-off condition have been combined into a self-consistent pattern in static magnetrons employing pure metal cathodes. Although cathode bombardment and flow of anode current are still unexplained, their existence can account for the observed enhanced emission, current division, and maximum current boundary. In this it is unnecessary to endow the cathodes with anomalous thermionic or secondary emission properties.

Partially complete empirical pictures now exist of the dependences of (1) the position of the maximum current boundary on secondary emission properties of the cathode surface, geometry, and magnetic field, (2) the flow of anode current in violation of the Hull cut-off condition on geometry and applied electric and magnetic fields, and (3) cathode bombardment on geometry and applied electric and magnetic fields.

The process causing cathode bombardment, enhanced emission, and flow of anode current in violation of the Hull cut-off condition is believed to be an electronic interaction which occurs almost entirely in the relatively dense region of space charge surrounding the cathode. The significant factors influencing the interaction appear to be the applied electric field, the applied magnetic field, the space charge density, and the thickness of the space charge sheath.

The basic problem of electronic interaction in the static magnetron has now been partially defined. It is expected that when a realistic solution to this problem exists, including a correct account of cathode bombardment, the occurrence of enhanced emission would be automatically predicted.

B. Significance of Enhanced Emission in Magnetron Design and Development

According to the usual magnetron theory (oscillating, multi-cavity magnetrons), efficiency should not fall off at low currents. Experimentally it always does. The large amount of cathode bombardment in the low current region can account, to a considerable extent, for this loss of efficiency. In the past attempts have been made to design magnetrons for low current operation. Such attempts have been uniformly unsuccessful, and will clearly continue to be so unless the *cause* of cathode bombardment is removed.

From present knowledge of the roles of thermionic and secondary emission in magnetron operation, it should be possible to predict the behavior of a given cathode in a particular magnetron. This is true if the thermionic and secondary emission properties of the cathode are unambiguously known. It cannot at this point be asserted that, for example, these properties are adequately known for oxide cathodes of the type used in many magnetrons.

Now that the magnetron's emission requirements are more fully understood, it should be possible in some cases to design cathodes suitable for particular requirements.

The authors wish to thank Professor W. E. Lamb, Jr., for guidance and many valuable discussions and suggestions. Appreciation is expressed also for the assistance and cooperation on many phases of the work by the staff of the Columbia Radiation Laboratory.

[12] A. V. Haeff, Proc. Inst. Radio Engrs. **37**, 4 (1949); Phys. Rev. **74**, 1532 (1948).
[13] J. R. Pierce and W. B. Hebenstreit, Bell System Tech. J. **28**, 33 (1949).
[14] O. Bunemann, C.V.D. Report, Ref. Mag. **37**, (1944); D. R. Hartree, C.V.D. Report, Ref. Mag. **31**, (1943); B. V. Rollin and K. W. H. Stevens, C.V.D. Report, Ref. Mag. **40**, (1944).
[15] R. Q. Twiss, "On the steady state and noise properties of linear and cylindrical magnetrons," thesis submitted for the degree of Doctor of Science at M.I.T. (1949).

Hollow Cathode for Positive Ion Studies in Cathode-Ray Tubes

C. H. Bachman, Harold Eubank,* and George Hall
Physics Department, Syracuse University, Syracuse, New York

(Received April 17, 1951)

Apparatus has been designed and tested which makes possible the identification of the positive ions in the beam of a cathode-ray tube. A hollow cathode construction allows the positive ions to pass through the conventional part of the electron gun and be studied as a positive ion beam by another set of focusing electrodes and deflection magnets. Curves of ion yield and bucking potential measurements indicate that most of the ions are formed in the early stages of electron acceleration.

THE supposedly pure electron beam of a cathode-ray tube actually contains positive and negative ions. The negative ions contribute to "burning" of the phosphor, this burning actually being a decreased fluorescent output rather than a physically visible discoloration. The positive ions travel toward the cathode where they cause emission troubles leading to short tube life.

The identities of the negative ions have been known for some time,[1] and various studies have been reported on this phase of the work.[2–4] The present paper is concerned with the building and testing of an electrode system suitable for the identification of the positive ions generated by a typical cathode-ray tube gun and screen arranged in a continuously pumped "demountable" vacuum system. It should be noted that the many sources of contamination present in a demountable system so becloud mass identity results that such results cannot be applied unreservedly to the case of carefully prepared sealed off cathode-ray tubes. Such a study will be the subject of a subsequent paper.

Normally in a simple electrostatically focused electron gun (Fig. 1) the positive ion effects are: first, a general blasting away of the oxide coating, and finally, in the advanced stages, a rather sharply defined hole drilled not only through the entire coating but through the metal cathode cap as well (Fig. 2). The positive ions apparently reach the cathode in a beam resulting from some focusing or collimating action. The beam diameter is commonly much smaller than the grid aperture.

This characteristic of the positive ion beam has been utilized in the present work. As the positive ion beam approaches the cathode through the grid aperture, it finds that a hole has been provided, and it passes through the cathode, emerging back of the electron gun as a beam which may be sent through another electrode system for focusing and deflection for mass studies.

This double-ended, hollow cathode gun is shown in Fig. 3. It will be seen that it consists of two electrostatic electron guns mounted back to back with suitable modifications in the cathode region. The cathode is formed of two concentric cylinders. At one end the ends are rolled toward each other and spot welded at the overlap. This rolled over portion is then covered with standard BaSr cathode coating. A tungsten heater is inserted in the annular space between the cylinders. Figure 4 shows the arrangement of the gun in the vacuum system together with the associated deflection and measuring devices.

The pumping system is conventional—a rotary pump

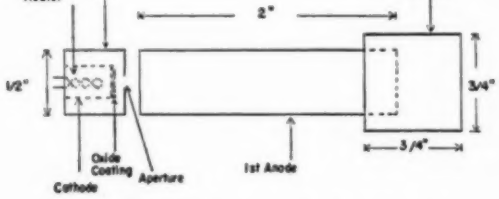


FIG. 1. Typical electrostatic electron gun.

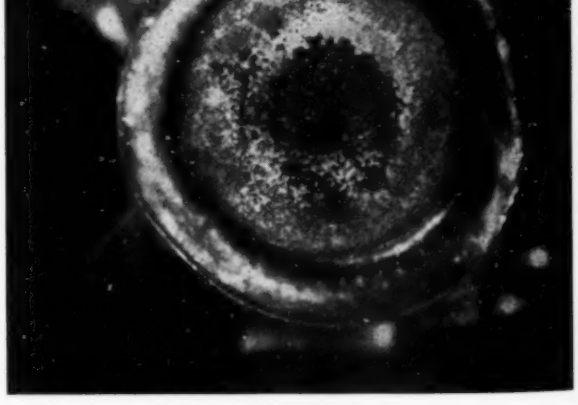


FIG. 2. Cathode destruction by positive ion bombardment. (Courtesy of General Electric Company.)

* Now at Physics Department, Brown University, Providence, Rhode Island.

[1] C. H. Bachman and C. W. Carnahan, Proc. Inst. Radio Engrs. **26**, 529 (1938).

[2] L. F. Broadway and A. F. Pearce, Proc. Phys. Soc. (London) **51**, 335 (1939).

[3] H. Schaefer and W. Walcher, Z. Physik **121**, 679 (1943).

[4] R. H. Sloane and C. S. Watt, Proc. Phys. Soc. (London) **61**, 217 (1948).

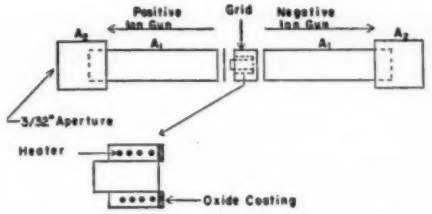


FIG. 3. Double-ended hollow cathode gun.

backing a two stage D.P.I. glass diffusion pump using Octoil-*S*. A trap is provided in the line so that use of refrigerants may be studied. All wax joints were made with Apiezon wax, and Celvacene was used on the Wilson seals. Vacuums never exceeded 10^{-5} mm Hg and, in general, were about 10^{-4} mm Hg.

The gun design permits accelerating potentials V_2 of from about one to ten thousand volts. Focused beams are obtained when A_1 is operated at V_1 values of approximately $\frac{1}{4}$ of V_2—this depending upon the screen distance. Electron current to the negative ion gun is controlled by placing a negative bias on G.

A simple electromagnet with 1-in. square pole pieces was used for deflection of the positive ions. Distortion, although noticeable at wide angles was not objectionable for the present work.

Ion currents were measured with a dc amplifier and a Leeds and Northrup galvanometer.

Initially it was not known whether the ions were formed all along the electron beam, at the fluorescent screen, or some gun aperture. In addition to the desirability of knowing the ion source, this information was also necessary for interpreting the deflection data and for focusing the positive ions on the screen. Such information was gained by determining through what potentials the ions had fallen when they arrived at the cathode. Before mounting the positive ion lens system, a metal collector of about 1 sq in. area was placed about an inch back of the cathode. Provision was made for rotating it and moving it about in space. This collector was connected, back to cathode, through a dc amplifier.

With zero grid bias and $V_1 = 270$ volts the ion current increased less than 5 percent as the second anode voltage was raised to 1000 volts. Although a slight dependence was indicated, it seemed evident that bombardment of the screen or other phenomena associated with the second anode potential played a relatively small part in ion formation. Most of the ions apparently

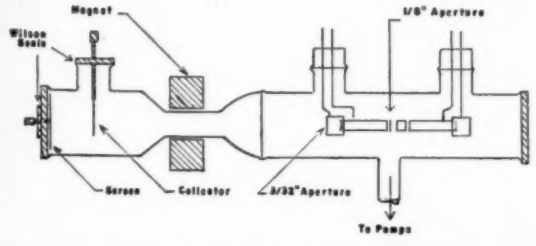


FIG. 4. Arrangement of apparatus for ion studies.

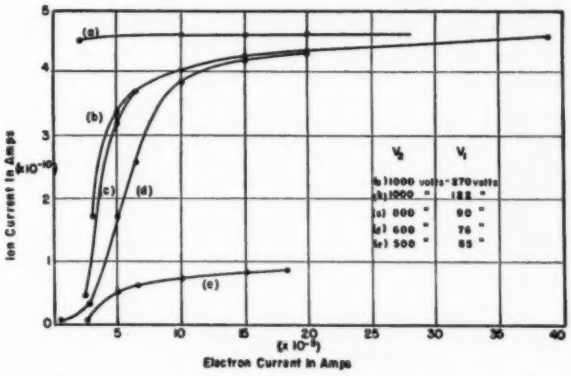


FIG. 5. Variation of positive ion current as a function of electron beam current.

are formed at or near the first anode. This was verified by subsequent stopping potential measurements.

The dependence of positive ion current upon electron current from the cathode is shown in Fig. 5 for several first anode potentials. Electron current was varied by means of grid bias. The same information is replotted in Fig. 6 to show the dependence of ion current upon first anode potential at various electron currents.

To determine the energies of the ion in the beam the collector and amplifier were arranged in a bucking circuit with batteries as a source of potential. Ion current plotted against positive bucking potential showed not a gradual or stepped decrease in ion current with increasing collector potential, but a sharp drop to zero at the critical voltage. The ions seem to have approximately the same energy. This energy seems to follow closely the first anode potential, as is shown in Fig. 7, where the critical potential is plotted against first anode potentials.

Physical movement of the collector probe indicated that the beam emerging from the cathode was highly divergent. This would seem to be quite probably caused by the magnetic field in this region resulting from the current of several amperes in the heater legs. Had the beam been found to have the relatively high second anode voltage of the electron gun and be sufficiently collimated, it would have been possible to let it move into the mass deflection field without further treatment.

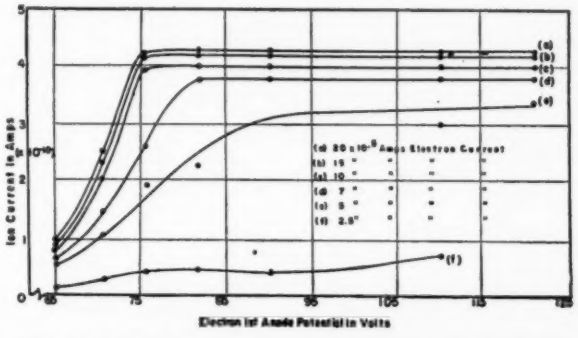


FIG. 6. Variation of positive ion current with electron energy.

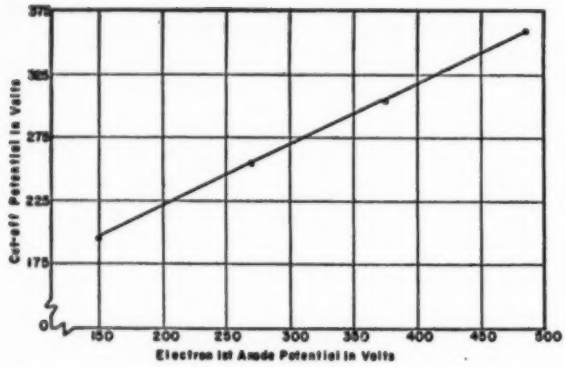


FIG. 7. Variation of maximum ion energy with electron energy.

Its divergence and low velocity made mandatory the accelerating and focusing system described earlier.

In the present work the intensities of the ion components usually was sufficient to excite a fluorescent screen, providing they were accelerated to a few thousand volts. The positive ion spectrum could thus be observed visually, and by noting the positions of the components on a scale scratched on the screen, measurements could be taken. The extreme sensitivity of fluorescent screens, particularly the sulfides, to ion bombardment is well known. Some of the negative ion workers earlier referred to found negative ion currents to be of the order of 10^{-13} amp. Positive ions have been found to be equally effective in screen burning. The "burning" effect sometimes occurs in a matter of minutes or even seconds to a degree that makes possible the use of the phenomena for recording purposes.

The fluorescent screens were prepared on a metal disk arranged to be rotated through a Wilson seal. The ions, deflected from the center, formed a spectrum consisting of a radial series of spots. When a given spectrum had been "burned" in, a new surface could be made available by rotating the disk slightly. The burned areas could be made visible by causing the entire screen to fluoresce. This could be accomplished by creating a discharge with a spark coil under increased pressure. For some measurements the disk was placed in a bell jar with a plane glass window, and photographs of the fluorescing screen were obtained for contact measuring. Since the axis of disk rotation usually didn't coincide with the undeflected beam, care had to be used in choosing the zero for each spectrum. Figure 8 is the photograph of such a series of spectra. Each radial row of spots is a positive ion spectrum.

FIG. 8. Photograph of fluorescing screen with numerous burned ion spectra.

Search for and measurements of the heavier masses is possible by increasing the deflection, the lighter components then moving out of view. In the case shown in Fig. 8 no other components were observed. It is quite probable that weaker components were present in the spectra and might be electrically measureable. That such background must be quite low, however, seems evident when it is remembered that the ion burning effect is cumulative, and relatively long exposures failed to develop additional burns. As this study is continued it is planned to obtain current measurements of the components, and the background will be investigated. At the present time interest is centered on the intense components which undoubtedly play the largest part in cathode destruction. In addition to its use in the demountable system described here, the hollow cathode gun construction is adaptable to sealed-off tubes in which the ion information may be directly correlated with life test data and production methods.

It is obvious that by adding a deflecting system to the electron-negative ion end of the double-ended gun, information concerning the positive and negative ion beams can be obtained simultaneously. This has been done successfully, and the work is continuing. A mass spectrograph capable of yielding the ions of both signs should have considerable value in fields other than that of cathode-ray tubes.

This work is in part sponsored by the Electronics Department of the General Electric Company. We are particularly grateful to V. C. Campbell for his interest and help in many ways.

The Power Spectrum of a Narrow Band Noise Passed through a Nonlinear Impedance Element*

JOHN L. MCLUCAS† AND RICHARD C. RAYMOND
Department of Physics, The Pennsylvania State College, State College, Pennsylvania
(Received April 19, 1951)

The power spectrum of the current flowing through a nonlinear impedance element excited by a narrow band of noise centered at 1.6 mc has been measured experimentally from the audio range through the sixth harmonic band. A noise having the spectral shape of an error function with a standard deviation of 3.8 kc was generated by passing the current from a photomultiplier tube through several tuned amplifier stages. The noise was applied through an attenuator of 50 ohms characteristic impedance to a type 1N34 crystal diode. The current flowing in the diode passed through attenuators of 50 ohms characteristic impedance to the spectrum analyzer. In the audio range, a conventional wave analyzer was used, and in the remaining ranges, a calibrated radio receiver was used to measure the spectral distribution of output power. The results are in qualitative agreement with the theory of Middleton. A possible reason for the poor agreement is the difficulty of expressing the diode characteristic in a simple form for the computations.

INTRODUCTION

MIDDLETON[1,2] and Rice[3] have developed methods of computing the power spectra of noise currents flowing in nonlinear impedance elements driven by noise voltages. The work reported here was undertaken to measure the power spectrum of noise current flowing in a type 1N34 crystal diode driven by a noise of known power spectrum. The power spectrum was also computed by substituting into Middleton's calculations a single term approximation of the measured characteristic of the diode.

EXPERIMENTAL EQUIPMENT

Figure 1 is a block diagram of the system used for the measurements. Noise from the noise source was attenuated to the desired level by attenuator number 1 and applied to the terminals of a power meter and the 1N34 diode. The noise currents flowing through attenuator number 2 were analyzed to determine the power spectrum. The continuous wave source was used alternatively with the noise source for calibration of the system. The attenuators were of 50 ohms characteristic impedance, and the power meters were of the thermal type with 600 ohms input impedance. The power spectrum of the input noise was found by substituting a linear element for the 1N34 and measuring the distribution of power with frequency with the analyzer. This step also served to check the harmonic generation in the analyzer and to show that the measured spectrum was actually that of the noise from the nonlinear element.

The noise source used was a type 931-A photomultiplier tube followed by a tuned amplifier for the selection of the band of noise desired. The 931-A has been shown to have a noise output which is essentially uniform with frequency in the range of interest. The tuned amplifier made use of 1.6 mc-tuned transformers for coupling through three stages of amplification. The output stage was a cathode follower with two type 6AG7 tubes in parallel, designed to operate linearly over a voltage range of three times the rms noise voltage used in the experiments. The noise amplifier thus operated without limiting for more than 99 percent of the noise amplitudes transmitted through it.

The attenuators used were T pad resistive attenuators inserted and removed from the line by means of push button switches and adjustable in steps of 1 db. They were rated by the manufacturer for frequencies up to 250 mc.

The continuous wave source used was a Hewlett-Packard Model 650-A test oscillator. This oscillator is provided with an output voltmeter and attenuator which can be used for calibration purposes. In this use, however, signals were taken from the oscillator in the

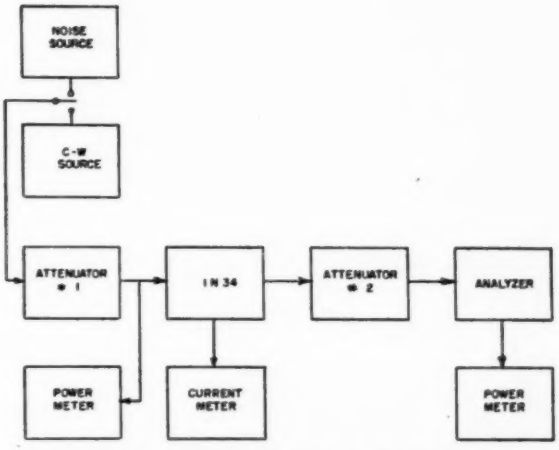


FIG. 1. System used in measuring the response of a type 1N34 crystal diode to a narrow band noise. The sine wave source was used for calibration of the analyzer.

* This paper is based on a thesis, "Generation of harmonics in nonlinear elements driven by noise," by John L. McLucas, submitted to the Graduate School of The Pennsylvania State College, 1950.

† Now with Haller, Raymond and Brown, Inc., State College, Pennsylvania.

[1] D. Middleton, J. Appl. Phys. **17**, 778 (1946).

[2] D. Middleton, Quart. Appl. Math. **5**, 445 (1948).

[3] S. O. Rice, Bell System Tech. J. **23**, 282 (1944) and **24**, 46 (1945).

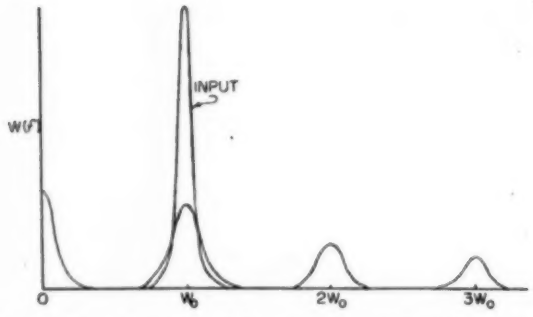


FIG. 2. General nature of the spectrum produced by passing a noise through a nonlinear element. Power levels and band widths have been exaggerated for purposes of illustration.

range from 0.5 to 1 volt, and the other attenuators were used for calibration of the analyzers.

EXPERIMENTAL PROCEDURE AND RESULTS

In an earlier paper, Middleton[1] has shown that the output of a nonlinear device which is driven by narrow-band noise is of the general form of Fig. 2. Here the output consists of "harmonics" of the input including the zeroth and first-order harmonics. In general, the width of the harmonic zones increases with increasing order of the harmonics. The powers in the harmonic bands are determined by input power and the transfer characteristic of the nonlinear element.

In the experimental work reported here, narrow band noise is generated by passing "white" noise through a series of tuned circuits. The resulting noise spectrum can be quite well represented by an expression of the form

$$W(f)=W_0 \exp[-(w-w_0)^2/w_b^2]. \qquad (1)$$

Here W_0 is the maximum value of the input spectral zone, w_0 is the center frequency of the input, and w_b is the departure from the center frequency at which the value of the function has fallen to $1/e$ of its maximum value.

In the experimental work reported here, the noise was of a much narrower band width than that shown in Fig. 2. The half-band width, w_b, for the input was 3.8 kc, while the center frequency, w_0, was 1600 kc.

Both the total power in each harmonic zone of the output and the spectral width of each zone were measured. The half-band widths of the various harmonics in the output are shown in Fig. 3. Here all the spectral orders have been normalized to unity peak value and shifted so that their center frequencies coincide on the graph; thus the "frequency" axis shows merely the departure from the center frequency for each harmonic band. The $m=1$ curve illustrates the spectrum of the output centered about the center frequency of the input. The actual width of this zone is very nearly equal to the measured width of the noise input. For all zones except the $m=0$ zone, the spectral width increases monotonically with increasing order of harmonic.

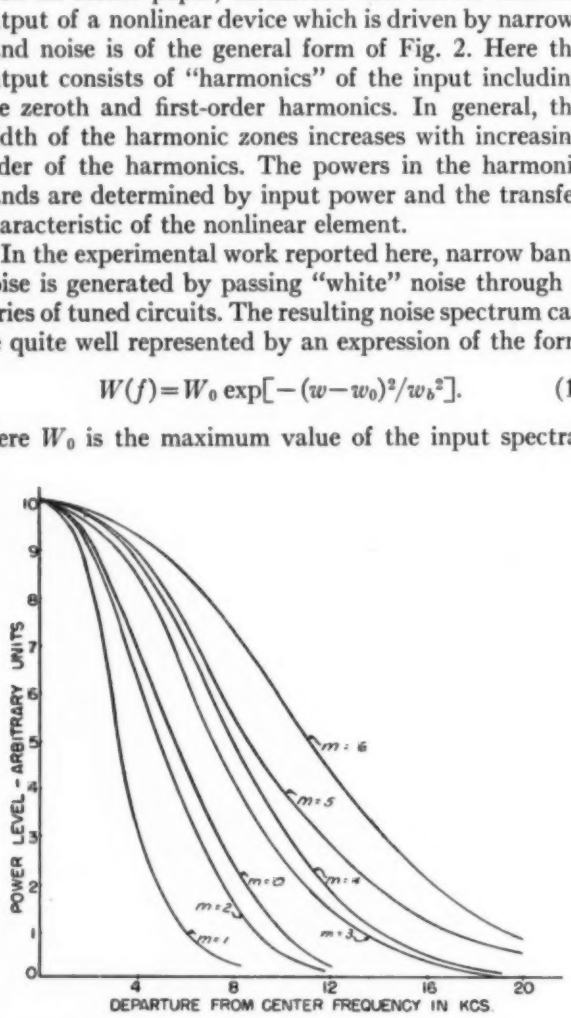


FIG. 3. Measured spectra of the harmonic components of narrow band noise passed through a type 1N34 crystal diode. The harmonic number is m. All spectra have been normalized to the same peak power for comparison of their bandwidths.

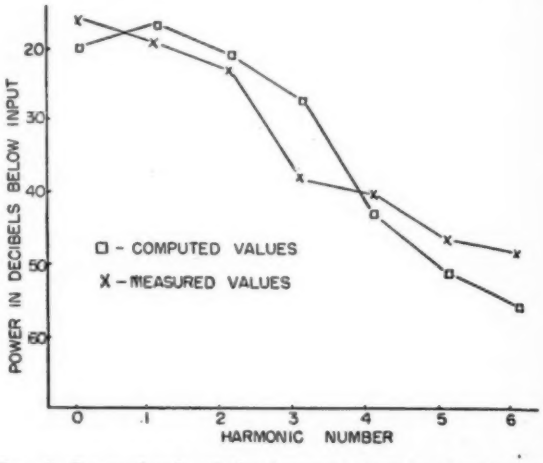


FIG. 4. Power levels of the harmonic bands of noise passed through a type 1N34 crystal diode and power levels computed by approximating the diode characteristic by a simple expression.

Figure 4 shows the amount of power associated with each harmonic band of the output. The output is measured in decibels below the input (0.6 volt rms).

COMPARISON OF RESULTS AND THEORETICAL PREDICTIONS

By the use of methods worked out by Middleton[2] we can calculate the output to be expected from a given nonlinear device excited by a given type of noise. The principal difficulty lies in representing the nonlinear transfer characteristic by a simple mathematical expression.

Since the method is well treated in Sec. 7 of reference 2, only the results will be given here.

To allow application of the simplest form of Middleton's methods, the nonlinear characteristic should be

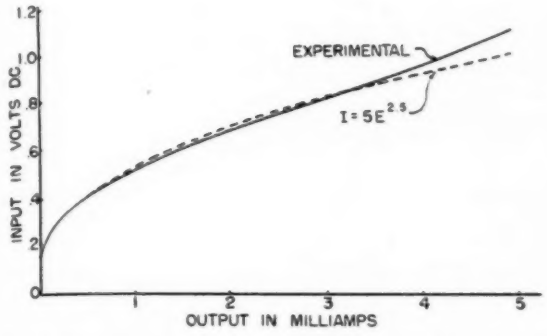


FIG. 5. Measured voltage-current characteristic for a type 1N34 crystal diode and approximate expression used in computation of the response to noise.

expressed in the form

$$I = BV^P, \ V \text{ positive},$$
$$I = 0, \ V \text{ negative}, \tag{2}$$

where I is output current and V is input voltage. The actual characteristic of the element used is shown in Fig. 5, marked "experimental." The curve which appears to be the best fit over the region of most interest is also shown in Fig. 5. Somewhat better fits could be obtained with power series representations including more terms, but it is desirable to shorten the calculations, which are long enough at best.

Middleton's method allows the correlation function of the output noise to be found in terms of the correlation function of the input noise, the power level of the input, and the characteristic of the nonlinear element. Since the details of the calculation are too lengthy to be reported here, only the substitutions will be given. The actual computations are reported in reference 3. The terms which must be known are:

ϕ	(average power input)	0.36 volt^2
B	(constant of Eq. (2))	5×10^{-3}
P	(exponent in Eq. (2))	2.5
$r_0(t)$	(the damping term of the normalized correlation function of the input)	$\exp(-w_b^2t^2/4)$
w_b	(half-band width of input)	3.8 kc.

The method used yields $R_m(t)$, the mth-order correlation function of the output, from the constants above. The power spectrum of the mth harmonic band of the output can then be found by taking the fourier transform of $R_m(t)$. Figure 6 compares the half-band widths for the various spectral zones with the measured values taken from Fig. 3.

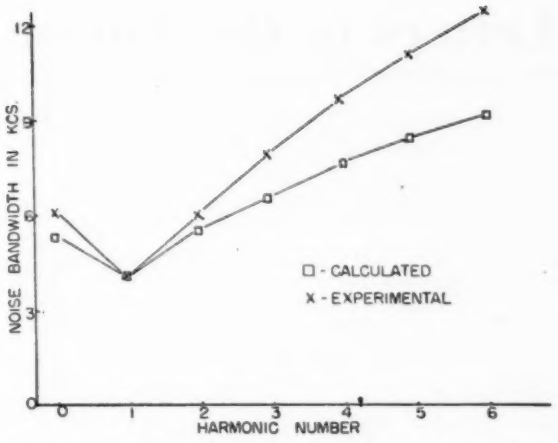


FIG. 6. Comparison of measured noise band widths with those computed on the basis of the characteristic of Fig. 5.

The power levels of the harmonic zones can be found directly from the $R(t)$'s by setting $t=0$ in these expressions:

$$R_m(0) = \int_0^\infty W_m(f)df. \tag{3}$$

This is just the power in any harmonic zone, m. The values computed are shown in Fig. 4.

For all spectral zones, $m=1$ to 6, the calculated band width is in each case $m^{\frac{1}{2}}$ times the width of the fundamental zone. The width of the $m=0$ zone is the same as $m=2$. In the experiment, w_0 was 3.8 kilocycles, and this is also the width for $m=1$.

As a matter of interest, the output power in each harmonic was measured for both noise and sinusoidal inputs of equal power. No significant differences were found.

CONCLUSIONS

The comparison of theoretical predictions and experimental data for narrow band noise through a nonlinear device shows that for this particular case, there is a general agreement between Middleton's theory and the experiment. The source of difficulty in applying the Middleton analysis is in describing the transfer function of the crystal rectifier by a single term of a power series. The theory does predict the general features of the measured spectrum. At the expense of complicating the analysis, a transfer function consisting of more terms of a power series could be used, but it is doubtful if the improvement in prediction accuracy would justify the additional effort involved.

Letters to the Editor

On Asymmetric Heat Conduction in Copper-Copper Oxide Rectifiers

F. HUBBARD HORN
General Electric Research Laboratory, Schenectady, New York
(Received June 18, 1951)

ASYMMETRIC heat conduction in a copper-copper oxide rectifier has been reported by Starr.[1] In spite of the difficulty of interpreting such an asymmetry, apparently no further investigation has been made.[2]

An experimental reinvestigation was made, using four different General Electric commercial copper-copper oxide rectifiers, type 6RC3 (K181), from which the "Aquadag" had been removed. Thermal conductance measurements were made, using a heat flow rather than a temperature-gradient method. The rectifier plates were not moved for conductance measurements of heat flow in reverse directions. Two rectifier plates were studied simultaneously in order to detect any technical differences in the apparatus. Measurements were made under three different conditions; *viz.*, with bare rectifier plates, with plates insulated on the copper face, and with plates electrically insulated but with the oxide face grounded. Within an experimental error of the order of 10 percent, no asymmetric heat conduction was observed in any case. The conductance of the bare rectifier was calculated to be 0.038 cal/°/sec/cm².

The discrepancy between these results and those reported by Starr[1] could be resolved if it were known that the thermocouples employed in Starr's work were not electrically insulated. A large Thomson potential is developed when a temperature gradient is imposed across a rectifier. This potential was measured and found to be of the order of 10 mv for a temperature difference of about 50°C. The sign of the potential depends on the direction of heat flow. Superposition of the Thomson potential on the thermocouple potential may have been possible in Starr's work because of the thermocouple circuits used. Such a superposition of the Thomson potential leads to an apparent, though erroneous, conclusion of asymmetric heat conduction when the temperature gradient method for measuring heat conduction is employed.

The conclusion that the copper-copper oxide rectifier is not asymmetric with respect to heat conduction is consistent with the conventional theoretical picture, since only a small number of electrons are involved in electrical conduction and their contribution to heat conduction is presumably obscured by lattice conduction.

I am grateful for the discussions with Dr. LeRoy Apker and Dr. Malcolm Hebb during this investigation.

[1] Chauncey Starr, Physics **7**, 14 (1936).
[2] H. K. Henisch, *Metal Rectifiers* (Clarendon Press, Oxford, 1949), p. 105.

Optical Absorption of Barium Oxide with Excess Barium*

EVAN O. KANE
Department of Physics, Cornell University, Ithaca, New York
(Received June 14, 1951)

SMALL crystals of barium oxide were grown by slowly cooling a solution of barium oxide in molten barium; this solubility has been established by Schriel.[1] The crystals are deep red in color and about 2×1×0.3 mm. They were grown in this laboratory by George Libowitz.

The crystal was held in a small absorption cell with glass windows to prevent the formation of the hydroxide, which would have occurred if the crystal had been exposed to ordinary air. The cell was mounted inside a vacuum system in good thermal contact with a container in which liquid air could be held. The absorption cell contained two apertures permitting the passage of light. One aperture was free, the other was covered by the crystal. The relative transmission curve of the crystal is given by the ratio of the light intensities through these two apertures. The transmission of the red barium oxide crystals was measured in the visible using a resolving power of about 200A. Data were taken for the crystal at room temperature and with liquid air in the container in contact with the crystal holder. No measurements were taken of the temperature, but from previous experience with the equipment it can be safely assumed that it was within the range −150°C to −190°C.

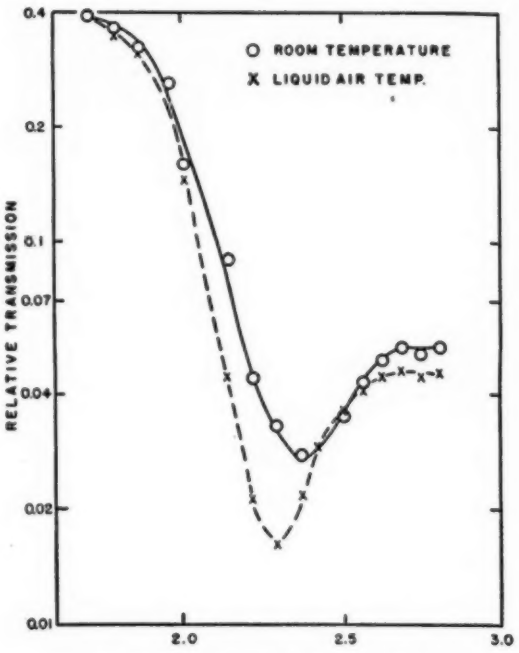


FIG. 1. Photon energy in electron volts.

Figure 1 shows two transmission curves for the same crystal, one for room temperature, the other for liquid air temperature. The lower temperature curve is sharper with a higher and some-

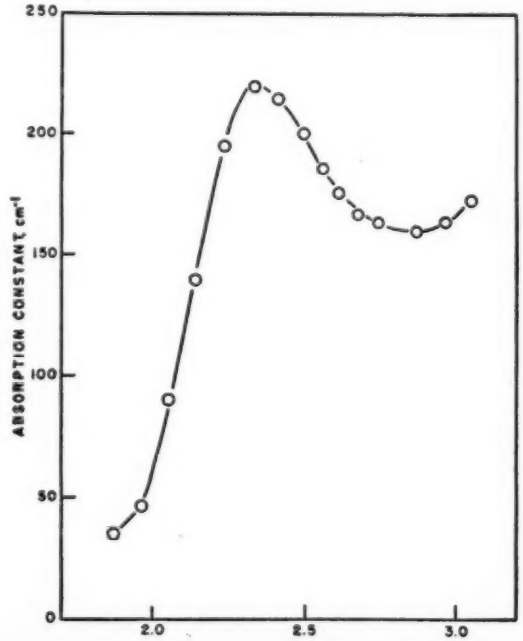


FIG. 2. Photon energy in electron volts.

what shifted peak. A possible explanation of the red coloration of these crystals is scattering from colloidal aggregates of excess barium. However, our measurements show a definite temperature sensitivity which would not be expected from colloidal scattering but which has a parallel in the *F*-center absorption of the alkali halides.[2] *F*-center absorption has been shown to be caused by electrons trapped at halogen vacancies. One way of forming *F*-centers is by the addition of atomically dispersed excess alkali metal. The wavelength and temperature dependence of the absorption is characteristic of *F*-centers but is also to be expected from any other atomically dispersed impurity centers. Our data gives no indication as to what type of center is formed. The simplest possibilities are interstitial barium atoms or oxygen vacancies which bind one or more of the valence electrons from the excess barium.

Figure 2 shows a plot of absorption constant *vs* photon energy in electron volts. The peak is evidently at about 2.4 ev. This is an optical activation energy. We expect the thermal activation energy to be much lower. This curve was calculated from the transmission curve of a crystal similar to the crystal for which the transmission curves of Fig. 1 were measured. On the assumption that the absorption is caused by atomically dispersed centers we can use this absorption curve to calculate the density of centers from a formula derived by Smakula.[3] Assuming an oscillator strength of unity we obtain 4×10^{18}/cc or 0.02 percent barium excess for the density of absorbing centers. Measurements are now in progress to determine the impurity concentration chemically. Preliminary results indicate that 0.1 percent barium excess is an upper bound.

* This work was sponsored by the ONR.

[1] M. Schriel, Z. anorg. u. allgem. Chem. **231**, 313 (1937).
[2] Frederick Seitz, Revs. Modern Phys. **18**, 384 (1946).
[3] Mott and Gurney, *Electronic Processes in Ionic Crystals*, p. 103.

Elastic Vibrations of Almost Cubic Parallelepipeds

H. Ekstein and T. Schiffman
Armour Research Foundation of Illinois Institute of Technology, Chicago, Illinois
(Received May 7, 1951)

THE frequencies of longitudinal vibration of a solid cube[1] exhibit several degeneracies, namely,

$$\rho\omega_1^2=2\mu k^2\{1+[(\pi^2+16)/\pi^2]\lambda/2\mu\},\quad \rho\omega_{2,3,4}^2=2\mu k^2,$$
$$\rho\omega_{5,6}^2=2\mu k^2\{1+[(\pi^2+4)/\pi^2]\lambda/2\mu\},\qquad (1)$$

where λ and μ are Lamé's constants, $k=\pi/\bar{a}$, and $\bar{a}$ is the edge length of the cube.

The six modes associated with Eq. (1) are taken as the zero order modes for the perturbation calculus. By distorting the cube slightly into a parallelepiped, these perturbation procedures lead to the familiar splitting of the degeneracies.

For the case of zero Poisson's ratio ($\lambda=0$), the sixfold degeneracy of Eq. (1) leads to a sixth-order secular determinant, whereas for all other values of Poisson's ratio, not too small, the perturbation calculus leads to one third-order, one second-order, and one first-order subdeterminant, analogous to the degeneracies in Eq. (1).

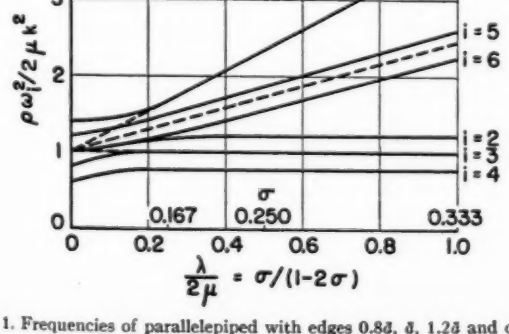


FIG. 1. Frequencies of parallelepiped with edges $0.8\bar{a}$, $\bar{a}$, $1.2\bar{a}$ and of cube with edges $\bar{a}$ as functions of Poisson's ratio.

Consider a solid parallelepiped with edges a, b, c and define an average edge as $\bar{a}=(abc)^{\frac{1}{3}}$, i.e., the edge of a cube whose volume is equal to that of the parallelepiped. The frequencies for this cube are given by Eq. (1). Let us further introduce the perturbation parameters $\epsilon=(\bar{a}-b)/\bar{a}\ll1$ and $\delta=(\bar{a}-c)/\bar{a}\ll1$. The frequencies of longitudinal vibration are thereby expressed in terms of $\bar{a}$, ϵ, δ, μ, and $\lambda/2\mu$:

Case I: $\lambda=0$.—

$$\rho\omega_1^2=2\mu k^2[1-2(\epsilon+\delta)];\quad \rho\omega_4^2=2\mu k^2[1-\delta];$$
$$\rho\omega_2^2=2\mu k^2[1+(\epsilon+\delta)];\quad \rho\omega_5^2=2\mu k^2[1+2\epsilon];\qquad (2)$$
$$\rho\omega_3^2=2\mu k^2[1-\epsilon];\quad \rho\omega_6^2=2\mu k^2[1+2\delta].$$

Case II: $\lambda\neq0$.—

$$\rho\omega_1^2=2\mu k^2\{1+[(\pi^2+16)/\pi^2]\lambda/2\mu\},$$
$$\rho\omega_{2,3,4}^2=2\mu k^2\left[1-\frac{2(\pi^4+8\pi^2+64)^{\frac{1}{2}}}{\sqrt{3}(\pi^2+4)}\Psi\cos\left(\frac{\phi}{3}+\frac{2\pi n}{3}\right)\right],\quad n=0,1,2,\qquad (3)$$
$$\rho\omega_{5,6}^2=2\mu k^2\left[1+\frac{\pi^2+4}{\pi^2}\frac{\lambda}{2\mu}\pm\frac{2}{\sqrt{3}}\left(\frac{\pi^2-2}{\pi^2+4}+\frac{\pi^2-8}{\pi^2}\frac{\lambda}{2\mu}\right)\Psi\right],$$

where

$$\Psi=(\epsilon^2+\delta^2+\epsilon\delta)^{\frac{1}{2}}$$

and

$$\phi=\cos^{-1}\left\{\frac{3\sqrt{3}(\pi^4+8\pi^3-128)(\pi^2+4)}{2(\pi^4+8\pi^2+64)^{\frac{1}{2}}}\frac{|\epsilon\delta(\epsilon+\delta)|}{\Psi^3}\right\}.$$

A group-theoretical analysis shows that the six modes of vibration associated with the frequencies (2) and (3) belong to the same one-dimensional irreducible representation A_1 of the $D_2(V)$ group, and hence in the frequency *vs* $\lambda/2\mu$ plane the values of Eq. (3) for $\lambda\neq0$, must be joined to those of Eq. (2) for $\lambda=0$, without any cross-over. The figure shows the eigenvalues for $\epsilon=-\delta=0.2$. Equation (3) has been assumed to be valid for $\lambda/2\mu>0.2$, and the curves joined without cross-over to the values given by Eq. (2). One accidental double degeneracy remains at $\lambda=0$; for all other values, all degeneracies, accidental as well as essential, are removed. An unexpected result is that the magnitude of the splitting is nearly constant for $\lambda/2\mu$ not too small.

[1] H. Ekstein and T. Schiffman, Phys. Rev. **77**, 757 (1950).

Surface Tension of Liquid Metals

N. R. Mukherjee
Department of Chemistry and Chemical Engineering, University of Washington, Seattle, Washington
(Received June 8, 1951)

IN a recent publication[1] it has been shown that the surface tension of liquid metals are very approximately inversely proportional to their respective molar volume. While working with a problem regarding the phenomenon of fusion, the author[2] has found that the surface tension of liquid metals at the melting point is also approximately proportional to their molar latent heat of fusion (Fig. 1). This correlation seems to be better than that of Atterton and Hoar,[1] although among the metals, whose surface tension has been correlated here, bismuth and antimony are exceptions. It should be pointed out that surface tension of metals are not known accurately and are difficult to determine. In many cases different values of surface tension have been reported in the literature. Moreover, surface tension of mercury has been reported to be anomalous and changes with time.[2] Therefore, it has been calculated by an equation derived by Mukherjee.[3] The empirical relation obtained from Fig. 1 is $\sigma=365Q_f$. where σ is the surface tension in dynes per centimeter and Q_f is the molar latent heat of fusion in kilocalories per gram atom.

The surface tension is believed to be due to the unbalanced forces of the surface atoms. For each surface atom, these unbal-

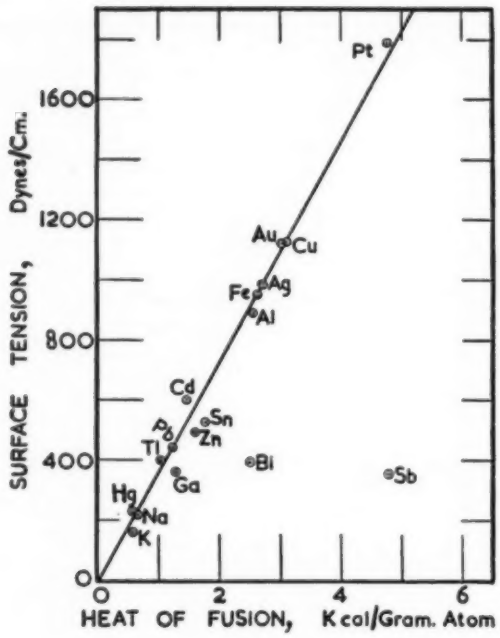


FIG. 1.

anced forces are a fraction of the total forces. Therefore, higher latent heat will be required for the fusion of a metal, whose atom has stronger forces, and, therefore, whose liquid has higher surface tension.

[1] D. V. Atterton and T. P. Hoar, Nature **167**, 602 (1951).
[2] R. S. Burdon and M. L. Oliphant, Trans. Faraday Soc. **26**, 205 (1927).
[3] N. R. Mukherjee, J. Chem. Phys. **19**, 502 (1951).

On the Orientations of Zinc Crystals Produced by the Bridgman Method*

L. M. SLIFKIN
Department of Physics, University of Illinois, Urbana, Illinois
(Received June 18, 1951)

IT has been observed[1] that for zinc single crystals grown by the method of lowering a crucible, sharply pointed at the lower end, through a vertical temperature gradient, the angle α between the c axis and the specimen axis was not random but was usually quite large, greater than 65°. Jillson,[2] however, using molds with blunt bottoms and a temperature gradient with a lateral as well as vertical component, obtained a random distribution of orientations. He suggested that preferred orientations might be the result of small amounts of impurities in the metal of other workers and that zinc of 99.999 percent purity would give a random distribution. Since the crystals of Slifkin and Kauzmann[1] were prepared from zinc of the same purity as those of Jillson and yet were usually of large α, it appears that the unique observation of Jillson may rather be due to some other factor. The possibility that the crucible design may be of importance in this respect immediately suggests itself. It is the purpose of this note to describe an experiment which supports this idea.

A number of zinc crystals were grown in glass tubes, the lower end of each pulled into a capillary and bent into the shape of a flat-bottomed fishhook, thus resembling an inverted, flattened question mark. On lowering one of these crucibles through a furnace, nucleation of the zinc crystal would be expected to take place somewhere in the horizontal section of the capillary hook. Some crystals were grown of 99.99 percent zinc, the rest of 99.999 percent zinc, kindly donated by the New Jersey Zinc Company.

The results are presented in Table I. The angle α, that between the c axis of the crystal and the direction of the unbent length of the crucible (i.e., the main body of the glass tube), is given only

TABLE I.

Crystal No.	Grade of zinc	α	θ
1	99.99%	61°	7°
2	99.99%	4°	8°
3	99.99%	61°	6°
4	99.99%	12°	11°
5	99.999%	39°	27°
6	99.999%	13°	13°
7	99.999%	3°	3°
8	99.999%	8°	6°
9	99.999%	45°	4°
10	99.999%	6°	6°
11	99.999%	45°	13°

for the sake of completeness. The important result is that θ, the angle between the direction of the horizontal section of the capillary hook and the basal plane of the crystal, is very small in almost every case. In other words, the basal plane was almost always approximately parallel to the direction of the flat section of the capillary. If the values of θ were randomly distributed, one calculates the average value would be 30°. The fact that for 10 out of the 11 crystals this angle is less than 15° strongly emphasizes the great effect that the geometry of the crucible can have on the orientation of the crystal.

I wish to thank J. S. Koehler and J. W. Marx for stimulating discussions.

* This work assisted by the AEC.
[1] P. W. Bridgman, Proc. Am. Acad. Sci. **58**, 165 (1923); **60**, 305 (1925). E. I. Salkovitz, private communication. L. Slifkin and W. Kauzmann, to be published.
[2] D. C. Jillson, Trans. Am. Inst. Mining Engrs. **188**, 1005 (1950).

A Note on the Initial Excitation of Linear Systems

L. A. ZADEH
Columbia University, New York, New York
(Received June 18, 1951)

IN a recent paper[1] the writer gave an expression for the response of an initially excited linear system to a given input. The system was assumed to be described by a relation of the form

$$L(p;t)v(t)=K(p;t)u(t), \tag{1}$$

where $p=d/dt$, $L(p;t)$, and $K(p;t)$ are polynomials in p with time-dependent coefficients, $u(t)$ is a given input (starting at $t=0$), and $v(t)$ is the response to $u(t)$. It was shown that in terms of the initial values of $v(t)$, $u(t)$, and their derivatives at $t=0+$, the response of the system to $u(t)$ may be expressed in the following form:

$$v(t)=\mathcal{L}^{-1}\{H(s;t)U(s)\}+\mathcal{L}^{-1}\left\{L^{-1}(s;t)\left[\left(\frac{L(s;0)-L(p_0;0)}{s-p_0}\right)v-\left(\frac{K(s;0)-K(p_0;0)}{s-p_0}\right)u\right]\right\}, \tag{2}$$

where s is the complex frequency $s=\sigma+j\omega$; $U(s)$ = laplace transform of $u(t)$; $\mathcal{L}^{-1}$ = inverse laplace transform; $H(s;t)$ = system function; $L^{-1}(s;t)$ = inverse of $L(s;t)$; p_0 is the initial value operator, i.e., $p_0{}^n v=v^{(n)}(0+)$ = value of the nth derivative of $v(t)$ at $t=0+$.

In the above formulation of the problem the basic assumption is that the specified initial values are those of $v(t)$ and its derivatives at $t=0+$, i.e., at the instant immediately following the application of the input. There are many practical situations, however, in which the initial values of $v(t)$ and its derivatives are specified at $t=0-$ (i.e., at the instant immediately *preceding* the application of the input) rather than at $t=0+$. The expression for $v(t)$ given by Eq. (20) is not directly applicable in such situations, since the constants $v^{(n)}(0+)$ involved in Eq. (2) are not known *a priori*. It is true that up to a certain n, $v^{(n)}(0+)=v^{(n)}(0-)$, but, in general, this equality would not hold for all of the $v^{(n)}(0+)$ in Eq. (2).

The case where the initial values of $v(t)$ and its derivatives are specified at $t=0-$, is not explicitly treated in reference 1. However, by using the general procedure described there, one can readily obtain a slightly modified expression for $v(t)$ which applies

to the case in question. Specifically, replacing 0+ by 0− and, in particular, extending the range of integration in the laplace transform to include 0−,

$$\mathfrak{L}\{v(t)\}=\int_{0-}^{\infty} v(t)e^{-st}dt, \tag{3}$$

the following expression for $v(t)$ is obtained:

$$v(t)=\mathfrak{L}^{-1}\{H(s;t)U(s)\}+\mathfrak{L}^{-1}\left\{L^{-1}(s;t)\left[\frac{L(s;0)-L(p_0;0)}{s-p_0}\right]v\right\}, \tag{4}$$

where $p_0{}^n v=v^{(n)}(0-)$, and all other symbols have the same significance as in Eq. (2). It will be noted that in this formulation the expression for $v(t)$ does not include the term involving the initial values of $u(t)$ (the input) and its derivatives.

For a fixed system defined by $L(p)v(t)=K(p)u(t)$, the result given above reduces to

$$v(t)=\mathfrak{L}^{-1}\left\{\frac{1}{L(s)}\left[K(s)U(s)+\left(\frac{L(s)-L(p_0)}{s-p_0}\right)v\right]\right\}. \tag{5}$$

It is seen that the expressions for the response on the 0− basis are similar to, and somewhat simpler than, the corresponding expressions on the 0+ basis.

[1] L. A. Zadeh, J. Appl. Phys. **22**, 782 (1951).

Errata: The Electric Tunnel Effect across Thin Insulator Films in Contacts

[J. Appl. Phys. 22, 569 (1951)]

RAGNAR HOLM
Stackpole Carbon Company, St. Marys, Pennsylvania

PAGE 572, Eq. (19), end of first line stands as (s/φ). This should be (s/eV).

Page 573, Eq. (30). The factor V is missing. The equation should read

$$\sigma=\frac{10^{14}V\exp[0.683(s/V)\{\varphi^{\frac{1}{2}}-(\varphi-V)^{\frac{1}{2}}\}]}{2.5nT^{\frac{1}{2}}(1-e^{-11600V/T})}.$$

Page 573. In the line before Eq. (31) stands as $V<0.5$. It should read $V<0.5\varphi$.

Figure 7 is correctly calculated according to the correct Eq. (30).

General Formulation for Calculation of Shaped-Beam Antennas*

A. S. DUNBAR
Stanford Research Institute, Stanford, California
(Received May 28, 1951)

THE synthesis of a shaped beam by the controlled variation of phase in an aperture having a known amplitude distribution may be accomplished by means of a method, stated by Chu,[1] based on the fundamental concept of the principle of the conservation of energy. In Chu's original method, which was applied to a cylindrical reflector and a line source, the shape of the reflector is calculated to produce the required shaped pattern from the given primary radiation pattern of the line source. This method has also been applied to the calculation of lenses and doubly curved reflectors to transform a given primary radiation pattern to the required shaped beam.[1,2] By means of the principle of stationary phase this method may be stated in general form.

To express this method in general form, we postulate an amplitude distribution $A(s)$ on a cylindrical surface, a plane curve of which must be slowly varying and restricted such that the curvature is always of the same sign. The far-field pattern is given by

$$E(\theta)=\int_{s_1}^{s_2} f(\theta,s)A(s)\exp[ik\phi(s,\theta)]ds, \tag{1}$$

where $\phi(s,\theta)$ is the phase function at the point P in the far field, and $f(\theta,s)$ is the form factor, which depends upon the polarization. Evaluation of Eq. (1) by the method of stationary phase yields the following results:

$$d\phi(s,\theta)/ds=0, \tag{2}$$

and

$$E(\theta)\doteq f(\theta,s_0)A(s_0)\exp[ik\phi(s_0,\theta)][-2\pi/\phi''(s_0,\theta)]^{\frac{1}{2}}, \tag{3}$$

in which s_0 is the point at which the phase is stationary, and ϕ'' denotes the second derivative with respect to s. Now Eq. (2) implies that

$$s_0=F(\theta); \tag{4}$$

hence,

$$\phi''(s_0,\theta)=(\partial^2\phi/\partial s\partial\theta)d\theta/ds_0. \tag{5}$$

Therefore, Eq. (3) becomes

$$E(\theta)\doteq f(\theta,s_0)A(s_0)\times\exp[ik\phi(s_0,\theta)][-2\pi/(\partial^2\phi/\partial s\partial\theta)(d\theta/ds_0)]^{\frac{1}{2}}. \tag{6}$$

Consequently, since $P(\theta)=|E(\theta)|^2$, we may write

$$-(\partial^2\phi/\partial s\partial\theta)P(\theta)d\theta=2\pi f^2(s,\theta)A^2(s)ds. \tag{7}$$

Equations (2) and (7) form the basis for the method of calculation to obtain an antenna which will produce a specified far-field pattern from a given amplitude distribution $A(s)$. Equation (7) relates a point on the antenna to a direction in the far field; and Eq. (2) provides a differential equation which must be solved for the required curve to produce the necessary phase function $\phi(s,\theta)$.

For the line source and cylindrical reflector, Eqs. (2) and (7) may readily be shown to yield, respectively,

$$d\rho/\rho d\alpha=\tan\tfrac{1}{2}(\alpha-\theta), \tag{2a}$$

$$P(\theta)d\theta=2\pi I(\alpha)d\alpha, \tag{7a}$$

in which $\rho=\rho(\alpha)$ is the radius vector of the reflector to be computed and $I(\alpha)$ is the (power) radiation pattern of the line source in the perpendicular plane.

Application of this general method to the calculation of progressive-phase antennas for shaped beams has been made with satisfactory results and successfully demonstrated experimentally by means of a wave-guide slot array.[3]

* This work was supported by Air Force Contract No. AF 19 (122) 78.
[1] L. J. Chu, *Microwave Beam-Shaping Antennas* (Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1947), Technical Report No. 40.
[2] S. Silver, *Microwave Antenna Theory and Design* (McGraw-Hill Book Company, Inc., New York, 1949), Chapter 13; A. S. Dunbar, Proc. Inst. Radio Engrs. **36**, 1289 (1948); D. W. Fry and F. K. Goward, *Aerials for Centimetre Wavelengths* (Cambridge University Press, London, 1950).
[3] A. S. Dunbar, "Progressive-phase shaped-beam antennas," paper presented at National Convention of Inst. Radio Engrs., March, 1951.

Errata: Application of the Radiation from Fast Electron Beams

[J. Appl. Phys. 22, 527 (1951)]

H. MOTZ
Microwave Laboratory, Stanford University, California

THE following formulas should be corrected as indicated below:

$$v_x=-e\mathcal{E}(1-\beta^2)^{\frac{1}{2}}\cos(\omega_0 t)/m_0\omega_0,$$

$$v_z=v(1-v_x{}^2/v^2)^{\frac{1}{2}}=\beta c(1-\tfrac{1}{2}v_x{}^2/v^2)=v\{1-[(1-\beta^2)e^2\mathcal{E}^2\cos^2(\omega_0 t)]/2\beta^2c^2\omega_0{}^2m_0{}^2\}. \tag{46}$$

$$z=v(1-a^2)t-[a^2\sin(2\omega_0 t)]/2\omega_0. \tag{47}$$

$$za^2/v+v(1-a^2)\alpha f(z)+\frac{a^2}{2\omega_0}\sin2\omega_0[(z/v)-\alpha f(z)]=0. \tag{49}$$

$$-\alpha f(z)=\frac{1}{v(1-a^2)}\left\{\frac{za^2}{v}+\frac{a^2}{2\omega_0}\sin\frac{2\omega_0 z}{v}\right\}\left\{1+\frac{a^2}{v}\cos2\omega_0\frac{z}{v}\right\}. \tag{50}$$

Three lines below Eq. (50): Eq. (57) should read Eq. (48). It should be pointed out that it has since been found possible to integrate the equations of motion exactly; no modifications of the rest of the paper are required. However, formula (73) should start with $(e^2/mc^2)^2$ and the second numerator is $\sin^3\theta\cos^2\phi$, while the exponents in the denominator are 5 and 3.

Announcements

Edgar Marburg Lecture Booklet

The 1950 ASTM Edgar Marburg Lecture by Dr. Wallace R. Brode, entitled *Chemical Spectroscopy*, has just been published in the form of a 52-page illustrated booklet. Copies of this booklet can be procured from American Society for Testing Materials, 916 Race Street, Philadelphia 3, Pennsylvania, at $1.35 each.

The purpose of the Edgar Marburg Lecture is to have described at the Annual Meetings of the American Society for Testing Materials, by leaders in their respective fields, outstanding developments in the promotion of knowledge of engineering materials. Established as a means of emphasizing the importance of promoting knowledge of materials, the Lecture honors and perpetuates the memory of Edgar Marburg, first Secretary of the Society.

A New Journal

We are introducing *Vacuum*, a new quarterly journal available on a subscription basis, published by W. Edwards & Company, Ltd., London, S.E. 26, England.

Vacuum is designed to fill a need in the scientific literature by reporting only on vacuum matters. It is addressed to the scientist and industrialist using vacuum procedures and further to specialists in all fields of science and industry to whom a knowledge of progress in vacuum matters is necessary or of value to their work.

Vacuum will consist primarily of the contributory section, of expert articles from international specialists and the abstracts extensively classified and specially printed for detached filing. Other feature series will be added as the need of the specialist readers becomes established.

Review of Naturalists' Directory

The Naturalists' Directory, thirty-sixth edition, has just been published at Salem, Massachusetts. The price of this Directory is $3.00 postpaid. It was first published in 1878 and since then has appeared every two or three years. It contains the names and addresses of hundreds of amateur and professional naturalists and a list of subjects of interest to them. It has been very helpful in bringing together naturalists in all parts of the world. In addition to the list of naturalists, there is a list of natural history museums and scientific periodicals. Naturalists generally have found the directory very helpful.

Conferences

The Sixth National Instrument Conference and Exhibit, sponsored by the Instrument Society of America, will take place in the Sam Houston Coliseum, in Houston, Texas, from September 10 through 14, 1951.

The fourth Conference on Gaseous Electronics, sponsored by the Division of Electron Physics of the American Physical Society, will meet October 4–6, in Schenectady, New York. Sessions will be held at the General Electric Research Laboratory in the new buildings at the Knolls, which were formally dedicated last autumn.

The 1951 National Electronics Conference will be held October 22, 23, and 24 at the Edgewater Beach Hotel in Chicago.

The conference is sponsored by the American Institute of Electrical Engineers, the Institute of Radio Engineers, Illinois Institute of Technology, Northwestern University, and the University of Illinois, with participation by the University of Wisconsin and the Society of Motion Picture and Television Engineers.